

TERMINOLOGY and LEXICOGRAPHY
RESEARCH and PRACTICE 23

Theoretical Perspectives on Terminology

Explaining terms, concepts
and specialized knowledge

EDITED BY

Pamela Faber

Marie-Claude L'Homme

John Benjamins Publishing Company

Theoretical Perspectives on Terminology

Terminology and Lexicography Research and Practice (TLRP)

ISSN 1388-8455

Terminology and Lexicography Research and Practice aims to provide in-depth studies and background information pertaining to Lexicography and Terminology. General works include philosophical, historical, theoretical, computational and cognitive approaches. Other works focus on structures for purpose- and domain-specific compilation (LSP), dictionary design, and training. The series includes monographs, state-of-the-art volumes and course books in the English language.

For an overview of all books published in this series, please see benjamins.com/catalog/tlrp

Editors

Marie-Claude L'Homme
University of Montreal

Kyo Kageura
University of Tokyo

Volume 23

Theoretical Perspectives on Terminology
Explaining terms, concepts and specialized knowledge
Edited by Pamela Faber and Marie-Claude L'Homme

Theoretical Perspectives on Terminology

Explaining terms, concepts
and specialized knowledge

Edited by

Pamela Faber

University of Granada

Marie-Claude L'Homme

University of Montreal

John Benjamins Publishing Company

Amsterdam / Philadelphia



The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/tlrp.23

**Cataloging-in-Publication Data available from Library of Congress:
LCCN 2022006872 (PRINT) / 2022006873 (E-BOOK)**

ISBN 978 90 272 1106 4 (HB)

ISBN 978 90 272 5778 9 (E-BOOK)

© 2022 – John Benjamins B.V.

No part of this book may be reproduced in any form, by print, photoprint, microfilm, or any other means, without written permission from the publisher.

John Benjamins Publishing Company · <https://benjamins.com>

Table of contents

Theoretical perspectives on Terminology: An introduction <i>Pamela Faber and Marie-Claude L'Homme</i>	1

Part 1. General Theory of Terminology (and beyond)

CHAPTER 1

The reception of Wüster's General Theory of Terminology <i>John Humbley</i>	15

CHAPTER 2

General principles of Wüster's General Theory of Terminology <i>Danielle Candel</i>	37

Part 2. Knowledge-based terminology

CHAPTER 3

Conceptual relations: From the General Theory of Terminology to knowledge bases <i>Anita Nuopponen</i>	63

CHAPTER 4

Terminology and standards <i>Sue Ellen Wright</i>	87

CHAPTER 5

Concept management for Terminology: A Knowledge Engineering approach <i>Ingrid Meyer</i>	111

CHAPTER 6

Multidimensionality <i>Lynne Bowker</i>	127

CHAPTER 7

Terminology and ontologies <i>Elena Montiel-Ponsoda</i>	149

Part 3. Socioterminology and Cultural Terminology

CHAPTER 8

- Founding principles of Socioterminology 177
Valérie Delavigne and François Gaudin

CHAPTER 9

- Cultural Terminology: An introduction to theory and method 197
Marcel Diki-Kidiri

Part 4. Textual Terminology, Terminology and Lexical Semantics

CHAPTER 10

- Textual Terminology: Origins, principles, and new challenges 219
Anne Condamines and Aurélie Picton

CHAPTER 11

- Terminology and Lexical Semantics 237
Marie-Claude L'Homme

Part 5. Corpus-based Terminology

CHAPTER 12

- Text genres and Terminology 263
Mojca Pecman and Natalie Kübler

CHAPTER 13

- Knowledge patterns in corpora 291
Elizabeth Marshman

CHAPTER 14

- Terminology and distributional analysis of corpora 311
Ann Bertels

Part 6. Terminology and Cognitive Linguistics

CHAPTER 15

- Units of understanding in Sociocognitive Terminology studies 331
Rita Temmerman

CHAPTER 16

- Frame-based Terminology 353
Pamela Faber

CHAPTER 17	
Conceptual metaphors	377
<i>José Manuel Ureña Gómez-Moreno and Francisco José Ruiz de Mendoza Ibáñez</i>	
Part 7. Variation and equivalence	
CHAPTER 18	
Causes of terminological variation	399
<i>Judit Freixa</i>	
CHAPTER 19	
Diachronic variation	421
<i>Pascaline Dury</i>	
CHAPTER 20	
Cognitive approaches to the study of term variation	435
<i>Sabela Fernández-Silva</i>	
CHAPTER 21	
Terminological growth	457
<i>Kyo Kageura</i>	
CHAPTER 22	
Terminology and equivalence	477
<i>Pilar León-Araúz</i>	
References	503
Standards and resources cited in chapters	575
Biographical notes	583
Index	591

Theoretical perspectives on Terminology

An introduction

Pamela Faber and Marie-Claude L'Homme

University of Granada / Université de Montréal

Since being recognized as an independent discipline in the first half of the 20th Century, Terminology has evolved both theoretically and methodologically in many different directions. Like other disciplines, it has adapted to new requirements, been influenced by new theoretical approaches and insights, and now employs a wide variety of methods and tools for work, study, and research. From the very start, Terminology began its path of expansion and widened its scope far beyond what anyone had thought possible. Evidence of this is the modified and now more diversified perception of specialized language units, their cognitive and communicative potential, and their role in society.

Over the years, perspectives on Terminology have evolved considerably as the behavior of specialized knowledge units in different linguistic, communicative, and cultural contexts was placed under the microscope. Terms had previously been assumed to foster unambiguous specialized communication, and in many cases, this was true since designations of specialized concepts, terms, helped to clear the path toward standardization. However, the exclusive relationship between a concept and a term was often challenged, particularly when terms were studied within the context of specialized communication.

The analysis of terms in corpora of specialized texts also raised many questions that had no straightforward answer. Many studies (see the chapters in Part 7 of this volume) highlighted the fact that the same concept could have more than one designation, and that these designations were not totally synonymous. In fact, today terms are now acknowledged to have communicative value and an expressive power of their own insofar as they are often steeped in the culture and ideology of the text sender, and encode metaphors that have an impact on the understanding of a specialized domain.

However, during this period of evolution and change, there is one year that stands out as a turning point that transformed Terminology forever. This watershed moment was at the beginning of the 1990s. The period before 1990 is strongly

associated with the General Theory of Terminology (GTT) whereas the period after 1990 witnessed a host of new approaches to Terminology. Some of these approaches evolved around the GTT and adapted to changing methods and tools, whereas other approaches differed considerably from the GTT, which certain scholars had begun to refer to as *traditional terminology*. Many started to question the biunivocality principle and the standardization objectives that it was designed to fulfill. Others claimed that the main issue resided in the fact that the theoretical principles of the GTT were not in consonance with linguistic and textual reality, and that the GTT view of concepts was too inflexible.

For over a decade, the fracture between the GTT and certain post-1990 theoretical perspectives was so deep that the merits of the GTT were buried under its possible flaws. This state of affairs prompted various researchers to revisit the GTT, analyze it in greater depth, and rediscover its merits. The reception of the GTT within the context of new approaches was also studied, and it was found that despite its drawbacks, the GTT was a pioneering theory that still has many virtues. In this volume, the principles of the GTT are revisited by Danielle Candel and John Humbley who contextualize them within the historical context in which they appeared.

There is no simple answer to the question of why Terminology has become a subject of such passionate debate. Part of the explanation resides in the various applications for which terminology work is carried out as pointed out by many authors in this volume. Standardization seeks to define rules to foster unambiguous communication. Language planning, on the other hand, attempts to bring about changes in the corpus or status of languages and/or language varieties. In doing so, it must take into consideration sociolinguistic and cultural factors.

Specialized translation is another very important application of Terminology. In most universities, Terminology is taught in translation curricula. However, translators often point out that the focus on concepts and knowledge organization of many terminological resources does not allow them to address all the questions raised by the translation of specialized texts. Terminological resources need to account for the linguistic and textual behavior of terms in addition to providing information about the kind of knowledge they convey. This includes the shades of meaning conveyed by term variants in different languages that designate the same concept.

In addition to these well-known applications, Terminology also has a close relationship with Information Science, Lexicography, Natural Language Processing, and Knowledge Representation. These applications, in addition to standardization, language planning, and specialized translation, raise different questions which often require answers that can only come from more diversified theoretical frameworks and methods.

Terminologists have an important role in the compilation of knowledge resources that are crucial to most of the applications that have just been listed. The design and development of specialized dictionaries, terms banks, ontologies, and knowledge bases contribute to the enhanced transmission and acquisition of specialized information. This requires a great many previous decisions concerning their content and structure.

However, the diversification of applications and resources only partly explains the profound changes that Terminology has undergone. From a theoretical perspective, there has also been considerable diversification and expansion. The foundations of Terminology have been considerably strengthened by the insights of Sager (1990), who pointed out three complementary dimensions of Terminology, namely, the cognitive, linguistic, and communicative dimensions. In this same line, Cabré (2000, 2003) states that these three dimensions should be an integral part of a theory of Terminology and proposed that terms themselves should be viewed as polyhedrons. The contents of this volume reflect the fact that research focusing on multiple facets of Terminology enhances our understanding of terms, concepts, specialized corpora, and specialized knowledge resources.

The chapters in this volume are a rich tapestry of different views on how to define a concept, a terminological unit, and a terminological relation. They also address a wide range of topics and issues in Terminology research from different perspectives.

General content of the volume

Our main objective is to provide readers with an opportunity to familiarize themselves with many different approaches to Terminology from Wüster to more recent initiatives. We also think that the time has come to take a step back and look at what these approaches have taught us. The time has also come to see how they address different issues and suggest new solutions to further a better understanding of the complex phenomena that terminologists must handle.

Some of these approaches have been developed within the context of other languages. However, the twenty-two chapters are all written in English in order to better convey ideas and concepts to interested scholars throughout the world. This volume is thus an invaluable reference for those who wish to understand Terminology as well as its different approaches and applications in all their variety and complexity. The authors have different backgrounds (Communication, Linguistics, Translation, Computer Science, etc.) and come from four different continents (Africa, America, Asia, and Europe). However, they all share a common passion for terminology, specialized knowledge organization, and specialized knowledge units.

Of course, it is impossible to cover everything in a single volume, nor can we foresee future developments, which is why it will be necessary to publish a second volume. There were also colleagues that had to decline our invitation, since they had previous commitments and their agenda was not compatible with our deadlines. However, we are confident that the chapters in this volume and the references that authors make to other work will give readers an overview of important insights, theoretical principles and methodological approaches that have been proposed over the years. The extensive list of references offers readers various ways to further explore specific topics.

The twenty-two chapters are organized semi-thematically and semi-chronologically. In our opinion, this organization will not only provide readers with a clear perspective on the evolution of ideas in Terminology, but will also give them an understanding as to why some of these ideas, which had long been taken for granted, were challenged. The only exception to this general organization is Part 7 in which we grouped chapters on variation and equivalence. Variation (and to a lesser extent equivalence) is a phenomenon that has become an integral part of terminological analysis regardless of the approach. However, here again, perspectives on variation can vary as reflected in the chapters in this part.

Part 1. General Theory of Terminology (and beyond)

The General Theory of Terminology has been subjected to much criticism, usually triggered by new applications that GTT did not specifically target. In a chapter on the reception of Wüster's General Theory of Terminology, John Humbley reveals that part of this criticism appears to be biased, since it may stem from a misapprehension of Wüster's works. He argues quite convincingly that Wüster's achievements can only be adequately assessed if they are considered within the historical context in which they were formulated.

In a chapter entitled "General principles of Wüster's General Theory of Terminology", Danielle Candel also reminds us of the importance of these principles. For the General Theory of Terminology (GTT), concepts and their designations are separate; conceptual analysis precedes the selection of suitable designations; and concepts must be clearly delineated in concept systems. Danielle Candel reviews Wüster's publications and correspondence and points out that Wüster foresaw many phenomena that are now being studied in much more recent work. In fact, to this day, the GTT has remained an important framework and its principles are applied, as such or in a modified version in many professional and international settings.

Part 2. Knowledge-based Terminology

Part 2 comprises chapters on approaches that are compatible with the General Theory of Terminology (GTT) to the extent that they focus on knowledge organization. However, they take a step forward and suggest ways to model complex phenomena, which include various types of conceptual relation. They also highlight challenges to classification and categorization, such as multidimensionality and the formal organization of knowledge.

Conceptual organization has the advantage of facilitating specialized knowledge acquisition. From a monolingual perspective, the specification of conceptual structures not only helps users to grasp the meaning of concepts but also shows their relations to other concepts in the context of a network. From a multilingual perspective, a conceptual structure is useful because it can act as an anchor point for linking terms in different languages, and provides a foundation for interlinguistic correspondence.

Work on knowledge modeling and representation allowed terminologists and other experts to account for relations other than the taxonomic and partitive relations that had been the focus of earlier work. This also led to the realization that concept systems could become extremely complex. Not only were there new sets of relations, such as cause-effect, function-object, etc., but relations were also subdivided into categories. For example, there are four or six subcategories of partitive relations, depending on the criteria taken into account. Anita Nuopponen's chapter on concept relations, "Conceptual relations: From the General Theory of Terminology to knowledge bases" shows how much the topic of concept relations has evolved and that constant revisions of typologies of relations are required.

Concept relations are important for standardization, which represents one of the first important applications of Terminology. The theoretical principles and assumptions of the GTT, especially the clear separation between concept and designation, the onomasiological approach and biunivocity, are undoubtedly useful to guide the choice of concept designations and the relations that link concepts. In her chapter on terminological standards, Sue Ellen Wright explains that standards can be applied to other important aspects of terminology work. They help define best practices for a wide variety of specialized resources and can be used in various settings, such as translation and knowledge management. Nevertheless, terminological resources have evolved and now come in a variety of forms and formats. Sue Ellen Wright offers a complete review of terminological standards, resources, and the different users targeted, while at the same time, situating them within the wider spectrum of language resources, such as terminological knowledge bases.

It was during the 1990s that researchers interested in knowledge modelling began to have access to sophisticated tools, such as terminological knowledge bases,

and ontologies, to store and represent concepts, which were linked to others by means of conceptual relations. They also benefited from the development of various technologies that help to encode, represent, and access terminological data.

In the early 1990s, Ingrid Meyer and her collaborators suggested that terminological data should be organized in a way that would make conceptual relations explicit. Terminological knowledge bases were thus born and have been referred to ever since by multiple authors in Terminology. For this reason, the volume includes the reprinted version of an article written by Ingrid Meyer in 1992, entitled "Concept management for Terminology: A Knowledge Engineering approach", which reminds us of how original her ideas were at the time and how important they have remained ever since.

Still another challenge raised by knowledge modeling is how to account for ways of classifying concepts. Lynne Bowker explains that concepts can be classified differently, depending on the characteristics taken into consideration. This inevitably results in knowledge structures that are less neatly organized, but which better reflect the different dimensions of concepts and their 'real' contribution to knowledge structures. She also explains that new approaches and tools now allow knowledge modelers to design complex concept systems.

Concepts, relations and multidimensionality can be represented in different kinds of resources, ranging from term banks that use standard data categories to very formal environments, such as ontologies. Terminologists and ontologists often work separately, but Elena Montiel-Ponsoda in her chapter entitled "Terminology and Ontologies" shows that they share many concerns, and could benefit from working together. She also explains that ontologies have become sophisticated environments that can now account for the linguistic properties of terms that denote concepts and concepts classes.

Part 3. Socioterminology and Cultural Terminology

Socioterminology and Cultural Terminology are approaches that arose in reaction to the General Theory of Terminology (GTT). Their proponents claimed that the GTT did not envisage the actual use of terms in professional, academic or social environments. Their fieldwork led them to question the biunivocal nature of terms or that each concept should have only one terminological designation. Many turned to frameworks and methods such as Sociolinguistics that implement a more descriptive approach. An excellent example of this kind of work is language planning where GTT principles were initially applied though in a modified and adapted form. Nations and official organizations dealing with language planning issues need to

implement and disseminate standardization rules, but must also take into account the sociolinguistic reality of the linguistic groups for which these rules are defined.

In a chapter titled “Founding principles of Socioterminology”, Valérie Delavigne and François Gaudin present one of the first approaches that positioned itself as an alternative to the GTT. They describe the context in which Socioterminology emerged, and the concepts that it defined. They also remind us of the important role of sociolinguistic principles and methods in Socioterminology. The authors explain some important concepts put forward by this approach, such as variation, which has become a central topic, and which is still addressed in more recent work.

Another important context in which terminology flourishes is culture. Culture provides the framework in which humans organize their thoughts, emotions, and behavior. In “Cultural Terminology: An introduction to theory and method”, Marcel Diki-Kidiri sets out the fundamental principles of the cultural approach to Terminology, where terms are studied as the cultural expression of specialized knowledge by a community of speakers. Cultural contexts thus reflect parameters of terminological variation, which represent the worldview of a certain community of speakers.

Part 4. Textual Terminology, Terminology and Lexical Semantics

Socioterminology opened the door to the study of terms in different contexts. Its emphasis on the professional settings in which terms are used soon drew the attention of researchers to the importance of term variation. The increased use of textual material to collect terms and information regarding them has enriched the representation of terms and opened new doors in Terminology research. In fact, the analysis of textual data reveals that knowledge is not expressed in a way that can be directly captured in a knowledge structure. Terms are subjected to phenomena that affect other linguistic units such as variation, polysemy, vagueness, etc. This has contributed to the creation of new paradigms for Terminology such as Textual Terminology, which highlights the need to take text genres into consideration.

Textual Terminology also aims to accurately portray how terms are used and how information given about them can be collected. The approach advocates the combined application of corpus linguistic methods, Natural Language Processing techniques and linguistic analysis to collect and interpret terminological data. In their chapter entitled “Textual Terminology: Origins, principles and new challenges”, Anne Condamines and Aurélie Picton explain the reasons why this approach was developed, and show that it is deeply rooted in applications. They also review important work carried out within the framework of Textual Terminology.

Knowledge representation and knowledge organization are crucial in many applications of Terminology. However, they often require methodological decisions that force terminologists to overlook the linguistic properties of terms. In her chapter entitled “Terminology and Lexical Semantics”, Marie-Claude L'Homme explores this topic and convincingly shows that viewing terms as lexical units or as items of knowledge leads to very different results, which include a larger variety of terms (nouns, verbs, adjectives, and adverbs), different structural patterns, and different relations between terms, among others. She argues that the lexicon-driven approach is more compatible with the behavior of terms in text and also recognizes that, though they are only partially compatible, both approaches are necessary in Terminology.

Part 5. Corpus-based Terminology

Assuming that knowledge is expressed in texts, many scholars have developed methods to locate and extract relevant pieces of knowledge from corpora. Methods and tools have been developed over the years that greatly facilitate these activities. These include large corpora, concordancers, term extractors, and knowledge extraction tools. However, the identification of terms in text and the retrieval of useful data is an extremely complex task since knowledge can be expressed in many different ways that are often difficult to predict. Part 5 contains chapters that address the expression, identification, and acquisition of knowledge from text. These chapters focus on specialized text genres, knowledge patterns, and distributional analysis applied to terminology and specialized domains.

In their chapter, “Text genres and Terminology”, Mojca Pecman and Natalie Kübler provide an overview of the studies, principles, and methods used to integrate genres in terminological analysis and practice. These authors define text genres as coded models according to which language users identify, interpret, and produce documents. In doing so, they show that specialized texts are more diversified than what was initially assumed and if this variety is overlooked, the result is excessively heterogeneous corpora. Their chapter also explains how Corpus-based Terminology and Textual Terminology can be applied to identify specific terminological features of genre types and to facilitate contrastive analysis of text genres, registers and domains.

With the growing popularity of corpus analysis in Terminology and the large quantity of specialized texts available, knowledge patterns (KPs) have also become a valuable tool for terminology work since they signal the presence of a certain type of semantic relation and the terms linked by it in corpus of texts. In “Knowledge patterns and corpora”, Elizabeth Marshman discusses the benefits and challenges of pattern-based approaches to Terminology. She describes how KPs have been

applied in previous research, the preparatory tasks required for their application, and the strategies used to evaluate them.

Over the last decade, Distributional Semantics has become very popular in Corpus Linguistics and Natural Language Processing for the extraction of semantic data from large corpora. In “Terminology and distributional analysis of corpora”, Anne Bertels surveys methods and models of distributional analysis in specialized domains and explores their potential for the analysis of specialized corpora, term extraction, ontology building, semantic analysis and term variation. Her overview of the research in the field shows how distributional analysis has been usefully applied to the creation of distributional thesauri, the extraction of semantic relations and knowledge patterns, and the identification of terminological variation.

Part 6. Terminology and Cognitive Linguistics

Recent cognitive-oriented perspectives on Terminology question the objectivist approach to concepts advocated by the General Theory of Terminology and argue that a rigid delineation of concepts is not compatible with the way concepts are represented in the human mind. The three chapters in Part 6 explain how terminology as well as its understanding, organization, and representation are related to cognitive processes such as understanding, frames, and metaphor.

In “Units of understanding in Sociocognitive Terminology studies”, Rita Temmerman describes the tenets of Sociocognitive Terminology in which the unit of understanding (UoU) is the core element. In her dynamic model of understanding, the process of understanding has: (i) a linguistic dimension since terms are understood in texts; (ii) a cognitive dimension since understanding is also categorization in the mind, and (iii) a communicative dimension since understanding is the outcome of communication and negotiation in a situated environment. A UoU is thus dynamic, prototypically structured, embedded in a sociological reality, and under constant negotiation.

Frame-based Terminology and its semantic, syntactic, and pragmatic microtheories are explained by Pamela Faber in her chapter “Frame-based Terminology” (FBT). Like Sociocognitive Terminology, FBT is also descriptive and text-driven. This approach applies the notion of ‘frame’ as an organized package of knowledge or cognitive context that relates elements and entities within a particular scene or situation. Frame-based Terminology specifically focuses on specialized knowledge representation, category organization and description, as well as the semantic and syntactic behavior of terminological units in one or various languages. Pamela Faber also shows how the principles of FBT are implemented in a terminological knowledge base, called EcoLexicon.

In “Conceptual metaphor in Terminology and specialized language”, José Manuel Ureña Gómez-Moreno and Francisco José Ruiz de Mendoza Ibañez provide an overview of the emergence of conceptual metaphor in specialized language, which shows that the embodied underpinnings of metaphor are indissolubly linked to sociocultural factors. These authors also highlight the pivotal role of metaphor, both as a tool to develop and communicate scientific and domain-specific thought as well as a resource for specialized knowledge popularization.

Part 7. Variation and equivalence

Finally, Part 7 includes chapters on variation and equivalence. Variation is an integral part of nearly all the approaches mentioned in the previous sections. Its importance in Terminology stems from the fact that it generated the first criticisms of the General Theory of Terminology (GTT), which did not contemplate term variation. This situation ultimately led terminologists to advocate more flexible approaches that were more in line with sociolinguistic and sociocultural reality. This section explains the different perspectives from which variation can be addressed and the possibility of establishing equivalence relationships between term variants in different languages.

In a chapter entitled “Causes of terminological variation”, Judit Freixa explains that variation is a natural and necessary part of specialized communication and that there are various reasons for the use of terminological variants. Her review of the different causes, based on an extensive body of literature, clearly demonstrates the importance that studies on variation have gained over the years, and is eloquent proof that variation remains a central topic in recent work. It also shows how much the topic has evolved since 2006 when a first article on this topic was published.

One of the causes of terminological variation is the evolution of knowledge, a fact that Terminology had overlooked until quite recently. Pascaline Dury, in her chapter “Diachronic variation”, explains that when terminologists only study terms from a synchronic perspective, they tend to ignore important phenomena such as the changes that can affect the delineation of concepts or the different names that are given to them. She presents pioneering work on diachrony in Terminology and shows that diachronic analysis contributes to a better understanding of concepts.

Sabela Fernández-Silva offers a different reading of variation which is grounded in a cognitive perspective. In her chapter, entitled “Cognitive approaches to the study of term variation”, in which variation is assumed to be cognitively motivated, she explores different cognitive factors that underlie variation in specialized communication. She also suggests that the analysis of variation would benefit from methods that combine corpus-based research and experimental studies.

Analysis of variation also includes research on the mechanisms at work when new terms are created. It is a well-known fact that new knowledge triggers a constant need for term creation. However, the rules underlying term formation and the factors at play are much more difficult to pinpoint. In his chapter entitled “Terminological growth”, Kyo Kageura suggests that linguistic patterns of term creation can be captured based on the combination of conceptual categories and conceptual systems.

In Terminology, equivalence has not received all the attention it deserves since it is often assumed to be an operation that consists in finding the designations in different languages that correspond to a given concept. In her chapter titled “Terminology and equivalence”, Pilar León-Araúz shows that there is considerably more to terminological equivalence than meets the eye. Various factors can interfere and act as obstacles to total equivalence. She also explains that equivalence must also be examined in connection with term variation as variation occurs in different languages. This means that true terminological equivalence is difficult, if not impossible, to establish without a detailed description of the nuances of meaning projected by term variants that designate the same concept.

Acknowledgements

The editors would like to thank all the authors who agreed to write a chapter on in-depth topics in a relatively short time. They also wish to thank Svetlana Komarova for her invaluable help in managing many editorial issues.

PART 1

General Theory of Terminology (and beyond)

The reception of Wüster's General Theory of Terminology

John Humbley
Université de Paris

Eugen Wüster is generally credited with founding modern Terminology theory. Over time, his legacy has been evaluated from various points of view, each new approach positioning itself against what came to be known as the General Theory of Terminology. Reappraisal came in particular from Socioterminology, Sociocognitive, Communicative and Textual Terminology and from Translation Studies. While some approaches set out to build on Wüster's principles, most sought to distance themselves from them. It is argued that his contribution can only be properly assessed by placing his activities in their historical context and by recognizing that theory was only one part of Wüster's achievement: he is at least as important for having embedded terminology in areas such as standardization and specialized communication.

Keywords: concept, discourse, onomasiology, reception theory, Socioterminology, standardization, synonymy, term, Translation Studies, transparency

1. Introduction

Eugen Wüster's role in founding Terminology theory is a subject of considerable controversy. There have been at least two theses devoted to this question (Campo 2012; Trojar 2017), one international conference (Candel, Samain and Savatovsky (eds.) forthcoming) and many articles, a selection of which will be reviewed here, in particular Antia (2001, 2002), Candel (2004, 2017), Humbley (2004, 2007), Myking (2001a, 2001b, 2005), Ilinca (2013). The most critical appraisals appear to have been made obliquely in studies concerned with breaking new ground in terminology research, in which the authors take pains to distinguish their own work from that initiated or inspired by Wüster.

This chapter attempts to give an account of the reception of Wüster's legacy from three complementary viewpoints: firstly and principally, the reappraisals made

in the context of new fields of Terminology (Socioterminology, Sociocognitive, Communicative and Textual Terminology and Terminology in the framework of Specialised Translation); secondly, the reception of one key element of Wüster's Terminology theory, the primacy of the concept; and finally, a note on some widespread misapprehensions concerning Wüster's thought, exemplified by the various understandings of the norm in Terminology. More generally, it will be argued that Wüster's heritage, all too often invoked by little more than hearsay, needs to be placed in its historic context, both academic and institutional, in order to ensure an objective appraisal of his contribution to what Terminology is today.

2. Wüster and the General Theory of Terminology

Eugen Wüster is generally credited with launching the first theory of Terminology,¹ which came to be known as the General Theory of Terminology (GTT) in circumstances explained in the second chapter of this volume (Candel, this volume; Budin 2006, 99). This formulation was used in particular by Wüster's successor, Helmut Felber,² whose article, subtitled 'In Memory of Eugen Wüster, Founder of the General Theory of Terminology', states namely that "the GTT is the theoretical background to the standardization of terminological principles" (Felber 1980, 67). It was also known as *The Vienna School of Terminology* (Felber 1981a, 70), the geographical reference suggesting not just academic credentials, in particular Wüster's involvement as honorary professor at the University of Vienna, but also alluding to the other terminology institutes which he was instrumental in founding either directly or through his successors. These include the secretariat of the ISO Technical committee (TC) 37 at the Austrian Standardisation Institute (ÖN at the time), the documentation centre Infoterm (founded in the 1971, first conference 1975), the TermNet network (1979/80) (Budin 2018, 96), TKE (Terminology and Knowledge Engineering, the conference of the Association for Terminology and Knowledge Transfer 1987), TAMA in Advanced Management Applications as from 1989, and Internationales Institut für Terminologieforschung (IITF, 1989). Highlighting 'Vienna' thus added important institutional kudos to emerging Terminology theory.

1. Rey (1979, 30, 2011, 906), Rondeau (1984), Lerat (1989, 51): 'founder of terminology and only theoretician', Gambier (1998, 49), Cabré (forthcoming). Dissenting voices were raised following his death: Sager (1990, 3) in particular denied any independent status for terminology as a discipline, another major subject of disagreement.

2. It is significant that the contemporary translation specialist Bühler (1982b, 429) calls Felber *a disciple* of Wüster's.

Wüster himself more usually referred quite simply to *Terminologielehre* (Budin 2018, 92): this was as he presented it as a new interdisciplinary field at the University of Vienna, then at the Copenhagen AILA conference in 1972 (Bühler 2005, 72; Budin 2018, 94)³ and as the title of his posthumous handbook (*Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, 1979). One of Wüster's main aims was to understand the role language plays in conceptualizing the world and in particular technology (Wüster 1959–1960) and to promote these considerations to the status of a science. Sager (1994a) in particular has been at pains to point out that the tautologous English rendering of *terminology science* is not an appropriate equivalent, and Rogers and Wright (2006, 108) plump for *terminology theory*, while acknowledging that Budin (2001, 20) modulates by referring to Terminology as several *theory clusters*.

3. Contextualising Wüster's contribution

Wüster's academic outlook was largely based on the intellectual climate before the Second World War. As the second chapter in this volume outlines, he was greatly influenced by the philosophical movement of the Vienna Circle, with its emphasis on Positivism and the Philosophy of Science (Nedobity 1984). This orientation was continued by close collaborators such as Oeser (1998); Felber (1984); Picht (2005) and after his death by philosopher/linguists such as Budin (2018, 90) and IT/linguists such as Schmitz (Budin 2018; Picht 2018).

Wüster is credited with having conceptualized Applied Linguistics (Weisgerber 1958) for the German-speaking countries (see Budin (2006, 96) for a more general appraisal), which is indeed announced as one of the main themes of his 1931 thesis. But Applied Linguistics as it was then had a very different agenda from what it has become today, i.e., a near exclusive specialisation in language acquisition didactics (Linn 2011), in particular in English-speaking countries. As Budin (2018, 91) makes it quite clear, Applied Linguistics at the outset entailed activities which would now be qualified as language planning (*Sprachgestaltung, Sprachregelung, Sprachnormung*). Even more recently, and in spite of some reorientation, Applied Linguistics is considered in the German-speaking world as being part of Linguistics

3. The article derived from the 1972 AILA paper (Bühler 1982b, 425) was published in a linguistics journal (*Linguistics*) as *Die allgemeine Terminologielehre: ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und Sachwissenschaft*, subtitled in English as *The Theory of General Terminology – An Intersection of Linguistics, Logic, Ontology, Information Science, and the Technical Sciences* (Wüster [1974] 1993) and variously translated into French as *Théorie générale de la terminologie* and *L'étude scientifique générale de la terminologie* (1976).

(Knapp 2011). And indeed, in spite of allusions to the contrary, Wüster kept up to date in his own reading of contemporary linguistic research: he quotes Saussure, Weisgerber, Trier and Whorf (Wüster [1974] 1993). In Wüster's view, Applied Linguistics was "language in its relationship to Man's activities in all spheres of life" (Antia 2002, 104). Ilinca (2013, 190) notes a shift in Wüster's thinking over the years, as he more latterly considered that Terminology was indeed an autonomous discipline.

This activity, carried out over many years, shows that Wüster was not only aware of current research in Linguistics, in particular the most advanced Linguistics of the period in contradistinction to Philology and he actively shaped the emerging field of Applied Linguistics, contributing to the broad scope of this field embodied in the numerous chapters of the German Applied Linguistics Association (Gesellschaft für angewandte Linguistik). Terminology was thus simply part of an overall orientation on language and the world. Wüster was thus a key player in Applied Linguistics as understood at the time in German-speaking countries.

His place in the field of Language for Special Purposes or LSP is very similar. He is recognized as a forerunner of LSP studies in 'Germanistik' (Zhu 2007, 251–255) and thus one of the founders of its broad-based epistemology. This goes some way to explaining the important hiatus which came into being between *Fachsprache* and the narrow didactic scope of what has become mainstream LSP in English-speaking countries. One can measure the surprise of a confirmed English for Special Purposes scholar, Shaw (2006, 31), when he found himself in a Terminology section of an LSP conference in 2005: rather than simply concentrating on teaching English to scientists, the focus was on how language, in relation with other codes, could express specialized knowledge.⁴ It was by no accident that this workshop had been organized by Heribert Picht, one of Wüster's principal successors.

German linguists such as Baumann (2005) confirm the broad nature of LSP seen from this viewpoint, highlighting the practical and theoretical functions of specialized communication, the role of convention, norms and codification, inspired from the Prague School of Linguistics. The central nature of LSP is such that Terminology manuals written in the wake of Wüster's work, such as the pioneering handbook by Picht and Draskau (1985), place Terminology within the framework of LSP, though by 2000, phraseology specialist Christopher Gledhill (2000, 22) interprets this as equating the terminological (i.e., conceptual) system with LSP.

4. ESP as I describe it here has grown out of the teaching of commercial and technical English, and retains an orientation less to the thought-patterns encoded in language choices than to the purposes and effects of texts, as in the journals *English for Specific Purposes* and *Anglais de spécialité* (Shaw 2006, 39).

In reviewing academic disciplines – Philosophy, Applied Linguistics and LSP – which have changed significantly since Wüster's time, care needs to be taken to avoid omitting other achievements which were instrumental in promoting his thought. In a private communication, Budin admitted that Wüster was important not just as a linguist, but as someone who got things done. It was Wüster who convinced the International Standards Association (ISA), the predecessor of the present-day International Standards Organisation (ISO), of the necessity to set up a technical committee to lay down terminology methods for standardization, a measure which has been incredibly successful. TC 37 was founded in 1936 and became ISO/TC 37 in 1951 as it is to the present day (Lang 1998, 625; Budin 2006, 98). This is a real Applied Linguistics success story.

In 1951 Wüster was commissioned by the European Economic Community to elaborate a pilot project for a technical dictionary based on ISO principles (Lang 1998), which eventually became *The Machine Tool* (Wüster 1968), hailed by Rey (2011, 906) as the embodiment of the terminological dictionary, but not before laying the foundation stones for Eurodicautom (Goffin 1997). Wüster was present at the foundation of the German Bundesprachenamt, where terminology databanks were first used in conjunction with Specialised Translation (Bühler 1982b, 428). He worked in close cooperation with UNESCO from its inception to establish documentation resources in the form of publications and documentation centres.

Documentation was a particular point of interest to him. He was a convinced supporter of the Universal Decimal Classification (UDC) (the terms in *The Machine-Tool* (Wüster 1968) are classified using this system) and actively collaborated with the International Federation of Documentation (FID) (Galinski 2019, 26–28) on both dictionaries and classification schemes, all of which eventually led up to the founding of Infoterm (Wüster 1974). With all these institutional achievements, plus a thesis in Linguistics, it is not surprising that the University of Vienna offered him an honorary professorship, which gave him the credibility as an academic.

This could be one reason that it is Wüster who is credited with having founded Terminology as a subject rather than, say, Captain Paasch (Van Campenhout 1997), whose remarkable maritime dictionaries anticipated materially much of what Wüster and others theorized only much later. Paasch, however, left no institutions and no theoretical writings behind him: all that was left were the dictionaries. Wüster can thus be credited not only for establishing Terminology and Applied Linguistics, but also for providing the prototype of what an applied linguist – of his times – should be.

4. Wüster's reception in new fields of Terminology

Having briefly set Wüster and his approach to Terminology in the historic context, we can now review the reactions in that of the following period. In Campo's thesis (2012), based on reception theory (Lakatos 1978), a basic division is made between approving and hostile reactions to Wüster's thought, a thoroughly reasonable orientation, though one which Van Campenhoudt (2006, 4) distances himself from, preferring to characterize as 'different cultural habits' when it comes to expressing disagreement. Myking (2001a, 55) on the other hand suggests a morally-based tripartite division: moderate and loyal (Scandinavian school of which he is a representative); radical and subversive (sociocritical, represented by Temmerman); and radical and loyal (once again Scandinavian with a cognitive orientation such as Nuopponen (2018b)).

The terminologists who engaged in this controversy generally expressed their adherence or their opposition in no uncertain terms,⁵ so it is useful to reflect on why Wüster's theory generated so much passion. This aspect will be touched upon in the following survey, but at the same time an attempt will be made to show, beyond the attitudes displayed, how developments in Terminology theory and practice carried on or diverged from Wüster's thought.

Reactions to Wüster's thought during his lifetime have been less thoroughly studied than the period following his death, so the generally positive reception which Campo (2012, 92) records for the period before 1979 may not be fully representative. It is known that Soviet linguists appreciated his 1931 thesis and had it translated into Russian. The first reference to his work in a French-speaking context appears to have been by the linguist and translation expert, Georges Mounin (1963, 127), who not unsurprisingly pays tribute to Wüster's work in setting up terminology databases for translation at the European Economic Community

5. The most frequent putdown is perhaps the insistence on Wüster's status as an engineer (and by implication not a linguist): Condomines (2005, 36) describes him as 'an engineer who published a dictionary on the Machine-tool in the 1930s [sic]'. Smith, Cuester and Temmerman (2005, 647) belittle him even more directly as a 'Viennese saw-manufacturer, professor of wood-working machinery, and devotee of Esperanto'. Wüster was in fact qualified as lecturer in wood-working in 1955 (Bühler 2005, 73). Examples of abusive language run from the psychological ('fetishistic obsession' Gambier 1991a, 42), through political ('linguistic fascism', *ibid*), medical ('blindness towards the interpenetration of the sciences' (*ibid*)), to a characterized dogmatism 'outdated positivism' (Humbley 2004). The Vienna School was seen to attempt to exert undue control in order to quash dissent: 'Centralized control over the theory of Terminology in order to avoid dissenting voices which might jeopardize its recognition as a theory', (Cabr  forthcoming, 34). Fighting words, indeed. There were however more flattering portraits, for example, Clas (2001, 582–583).

level, as mentioned previously. Within the German-speaking academic world, his earliest advocate was Leo Weisgerber, who arranged for a special issue of the journal *Sprachforum – Zeitschrift für angewandte Sprachwissenschaft* to mark Wüster's 65th birthday, foreshadowing the posthumous *Festschrift* (Felber, Lang and Wersig 1979): *Terminologie als angewandte Sprachwissenschaft*. Both titles significantly place Terminology in the field of Applied Linguistics. This work was carried on by Heribert Picht, who had been a guest scholar at Wüster's home and business in Wieselburg in 1973 (Budin 2018, 95). Picht subsequently created Terminology courses at the Copenhagen Business School, branching out into Scandinavian co-operation with Laurén in Finland, Myking in Norway, laying the foundations of the still very active Nordterm network. Similarly, terminology networking prospered in Germany thanks to the impetus of Schmitz among others.

Outside the German-speaking world, Wüster sprang to prominence in the wake of the 1960s *Révolution tranquille* in Quebec. The Office de la Langue Française (OLF)⁶ was entrusted with ensuring the status of French as the working language in organisations and businesses according to the mandate given by the Charter of the French Language, adopted in 1971. It was in this context that the OLF actively consulted language experts and organisations in the form of a series of conferences held between 1971 and 1977 (Corbeil 2003, 264) to establish a theoretical groundwork to support the OLF's practices and to better distinguish Terminology from Translation. It was to this end that Wüster was invited in 1975 to the Lac-Delage conference to give a paper entitled *Essai de définition de la terminologie*, a shortened version of his 1972 AILA paper (Wüster 1976), which was incorporated into a series of books edited by the Université Laval.

Most of the critical reactions to Wüster's works, however, came after his death in 1979. As far as Terminology reference books are concerned, the major event was the posthumous publication of Wüster's handbook, which ran to two reeditions (Humbley 2007). It was followed up by the UNESCO sponsored manual by Wüster's close associate, Helmut Felber, first in an English version (1984), then in a Spanish version with Picht (1984), and finally a French (1987) translation. This international handbook is sometimes thought of as radicalizing Wüster's viewpoints (Cabr   forthcoming, 14). Wüster's work was actively followed up in particular in Denmark, where the second edition of his handbook was published in 1985 and where the mission of the Copenhagen Business School (previously Handelsh  jskolen i K  benhavn) was to apply these terminological methods to fields not envisaged by Wüster himself, in particular law, economy and the social sciences (Picht 1981; Arntz 1985, 44).

6. Founded in 1961, this organisation has changed its name over the years. It became the *Office qu  b  cois de la langue fran  aise* in 2002.

At the same time, textbooks in Terminology were appearing in Spanish (Felber and Picht 1984; Fedor de Diego and Boroni di Sanchez-Vegas 1984; Fedor de Diego 1995) and in French, with an important Quebec input. The first was a practical handbook by Robert Dubuc (1978), who, although he had met Wüster (Cabr   1996, 230), does not mention him. The second was a university textbook first published in 1981 by a linguist active at the Universit   Laval, Guy Rondeau (1984), who had been working in cooperation with the (then) R  gie de la Langue Fran  aise (Office [qu  b  cois] de la langue fran  aise), which was already deeply engaged in terminology work aiming at ensuring French as the language of the workplace. W  ster's theories are given prominence as the main framework to understanding terminology. Rondeau is sceptical however about the possibility of drawing up watertight domains.

Terminology practice, known as terminography, got off to an early start in Quebec with Auger and Rousseau (1978) on methodology, subsequently adapted to Catalan, Spanish, and Portuguese, whereas groundwork had begun as early as 1971 (Compagnie G  n  rale   lectrique du Canada, Usine de Qu  bec) producing glossaries corresponding to attested needs and directly used on the assembly line. Meanwhile, in France, a volume was published by Alain Rey (1979) in the popular encyclopaedia series *Que sais-je?*, which situates Terminology from an epistemological point of view, acknowledging W  ster's founding work (Rey 2011, 902). Campo (2012) qualifies the immediate reaction to W  ster's work immediately following his death as overwhelmingly positive, probably as in the German-speaking world reactions came from scholars who had actually worked with him. This was even as early as the second edition of W  ster's manual in 1985. However, Picht felt obliged to deflect criticism already addressed to the theory (Humbley 2007, 86).

Since space precludes a complete review of subsequent developments in Terminology which led to a generally negative appraisal to W  ster's thought, the most representative fields are briefly described here, to the regrettable exclusion of Cultural Terminology (Diki-Kidiri 2000, this volume). Another area regarded as important to Terminology by W  ster and deserving of a more thorough treatment is Documentation, which has seldom been appraised by terminologists, with the notable exception of Holzem (1999, 5). Holzem, in spite of a socioterminological background, points out striking parallels between methods in Documentation and in Terminology. In addition, W  ster's interest in IT and Computer Science at the end of his life was partly inspired by the facilities computers offered for terminological information and documentation (Budin 2018, 94).

4.1 Socioterminology

The first organised salvo against principles of the General Theory of Terminology came from France and the emerging field of Sociolinguistics applied to Terminology (Delavigne and Gaudin, this volume). A highly fruitful collaboration had taken place from the early 1970s (i.e., post May 68) at the new and innovative university at Nanterre where Louis Guilbert, the foremost representative of historical specialised vocabulary in French, worked with a number of sociolinguists, some of whom were specialized in language inequality and others in the use of language in the workplace (Marcellesi and Gardin 1974). Many of these moved on to the University of Rouen, where courses in Terminology with an original sociolinguistic background were organised under the direction of Louis Guespin.

It was a young scholar from this background, Yves Gambier, already involved in Translation Studies in Finland, who led the charge at a conference on cross-fertilisation in Terminology. His paper on acid rain (Gambier 1987) attacked several of the basic principles of Wüsterian Terminology, outlined below. It was followed up by a series of publications, notably a special issue of the Rouen-based *Cahiers de linguistique sociale* (Gaudin and Assal 1991), and subsequently by Brussels-based journal *Le Langage et l'homme* (Gambier and Gaudin 1993). Gambier was soon joined by François Gaudin, who defended the first thesis in Socioterminology, supervised by Louis Guespin, which gave rise to two substantial monographs (Gaudin 1993, 2003) and a useful synthesis (Gaudin 2005). In all of these, the socioterminological orientation is portrayed as the antithesis of Wüsterian Terminology theory.

One bone of contention was the issue of term transparency: it was claimed by Wüster's detractors that terms were not necessarily motivated so as they fit into a predetermined conceptual pattern; one-to-one correspondence or biunivocity was neither observable nor desirable: the idea of monosemy cast into doubt that the 'ideal' term is motivated, systematic, can be derived from and shortened (Gambier 1987, 319). Although this paper appears to be a frontal attack on a certain conception of Terminology, neither Wüster nor the General Theory of Terminology is mentioned explicitly. The object of the attack is simply designated as 'la terminologie dominante', a formula to be widely used in papers by socioterminologists. A comparison of the points of conflict could be ISO 704 standard for term formation (see Table 1) or alternatively the neology handbook produced by the Conseil International de la Langue Française (Diki-Kidiri, Joly and Murcia 1981), which was actively involved in official French language planning activities, so the criticism could also be aimed at official terminology.

Delavigne and Guespin (1992, 19) defined Socioterminology as a critique of the "conceptual tools of dominant terminology", and Gambier (1998, 49) later made

explicit the reference to Wüster. Even though ‘dominant’ Terminology has several sources, it is indeed Wüster “an engineer and an industrialist” who represents the postulates as they were formulated, which formed the foundations for methods of collecting and processing concepts and terms. However, no precise quotations from Wüster are given.

Table 1. Comparison between Gambier’s objections and ISO standard 704

Gambier’s (1987) points of fracture	ISO 704:2009 recommendations
Term is not necessarily motivated	A term is transparent when the concept it designates can be inferred.
Terms do not necessarily belong to a predetermined structure	The terminology of any subject field [...] a coherent terminological system corresponding to the concept system
There is not necessarily biunivocity between term and concept	Monosemy: ... a given term is attributed to one concept and a given concept is represented by only one term [...]
Derivability is not obligatory	Productive term formations that allow derivatives [...] should be favoured
Terms are not necessarily short	A term shall be as concise as possible

The sociolinguists’ dissatisfaction with the General Theory of Terminology stems from both a different linguistic environment and divergent aims. The most important difference in aims and environment was certainly the attitude towards discourse both as an object of study and a communication tool. The sociolinguists observed that in real specialized exchanges, variation was rife and the ideal of one term representing one concept was rarely attained and perhaps not even desirable. Working as a linguist who at the same time was an engineer, in collaboration with other engineers, the founding father of Terminology found that discourse was not particularly useful as a tool, though, as will be demonstrated later, actual usage was taken into account in Wüster’s method.

It has been claimed that sociolinguists worked in a monolingual environment (Lerat 2000). This is only partly true as a great deal of the Sociolinguistics output was centred around language minorisation and linguistic insecurity, in particular Corsican as opposed to French – Marcellesi’s main interest – or Arabic as opposed to French and English, and – more closely linked to Terminology – scientific French under the pressure of publications in English (Gaudin 1993). This opened a field of common interest with Quebec linguists, especially terminologists and lexicographers. Jean-Claude Boulanger, who, as a lexicographer, seldom worked directly in the field, is credited with coining the expression *Socioterminology* in 1981

(Boulanger 1995a, 197). This focalisation on combatting language inequality, which led to a whole series of studies as from the 1980s on terminological implementation (*implantation*) (Auger 1999), was however peripheral in Wüsterian terms. Wüster defined the role of terminology as a facilitator of international specialized communication, where the important factor was efficiency. In his thesis he compares the material advantage of Esperanto as a universal scientific language rather than English (Wüster [1931] 1970).

One particularly important offshoot of socioterminology work is represented by Leïla Messaoudi, working in the multilingual society of North Africa on 'ordinary technolects', insisting on the relevance of the context of oral exchanges in the workplace rather than the abstraction of concepts as Wüsterian theory aimed for (Messaoudi 2006, 687–697). Sociolinguists also rejected Wüster's distinction between word and term, considering that terms are simply words used in a particular way (Alexandru and Gaudin 2006, 59).

The final aspect of the socioterminologists' agenda to be touched on here, which features prominently in the monographs of Gaudin (1993, 129, 2003, 126) and Delavigne (1995), is the importance of terminology in popularizing specialised knowledge, or *vulgarisation* as it is known in French. For Wüster, the focus was communication between specialists and it was left to his successors to envisage how the general principles could be applied to communication between lay and expert.

4.2 Communicative and cognitive developments of Terminology

Like Socioterminology, the cognitive approach to Terminology was described as radical and subversive by both Myking (2001a) and Antia (2001). Like Socioterminology too, these new lines of research were being developed inspired by developments in Linguistics, in particular the communicative approach, theorized by and around Maria Teresa Cabré (2000, forthcoming) and the sociocognitive orientation pioneered by Rita Temmerman (1997, 2000). Cabré, already fully trained as a linguist, discovered Terminology on an extended stay in Quebec at the Office de la Langue Française (Cabré forthcoming, 154), and immediately recognized its relevance for a similar language policy, with Terminology at its centre, to promote Catalan as the language of the workplace in Catalonia, founding in her wake TermCat, the Catalan terminology unit (Cabré forthcoming, 4).

In both cases, status language planning entailed terminology development, thus the identity of interests. It was concerned with how terms are used socially that brought the realisation that terminology had three dimensions which needed to be simultaneously taken into account: the cognitive dimension, which Cabré recognized as adequately dealt with by Wüster, but also the linguistic and the

communicative dimensions as well. Her work at the Applied Linguistics Centre (IULA) at the Pompeu Fabra University enabled her not only to train a generation of linguist-terminologists, but also to launch several large-scale terminology and neology networks.

Cabré is close to Wüster inasmuch as their material, institutional and scholarly achievements effectively reinforced the overwhelming impact they have had on terminology studies. Contrary to Wüster, though, Cabré places Terminology firmly within the field of Linguistics and thereby heralds one of the foremost themes of the 1990s and beyond: the relations between Terminology and Linguistics (Myking 2020). Pamela Faber (2005, 110), shifting focus, remarked that it was Linguistics which had in fact ignored Terminology and she set about rectifying the situation for translators by basing terminological principles on Cognitive Linguistics using Frame Semantics.

Sociocognitive Terminology was also conceived out of dissatisfaction with what was felt as the inadequacies of Wüster's accounting for basic concepts of Terminology. Zawada and Swanepoel (1994, 253) were among the first to note the use by scientists of prototypical or scalar categories "embedded in experientialist epistemology" in such a concrete subject as Mineralogy rather than sharply drawn distinguishing features (Zawada and Swanepoel 1994, 261). Temmerman (2000, 27–28) paid particular attention to this article in a broadly-based review of how Terminology theory had developed since Wüster and confronted it with the reading she gave of the development of the terminology of life sciences, in particular DNA sequencing. By using conceptual linguistics, she was able not simply to account for new terms, but to explain how and why they were formed, relying on the rereading of metaphor theory. Once again, developments in Linguistics invited scholars to confront new attitudes towards terminology work. This notwithstanding, Trojar (2017, 66) points out correctly that Wüster never claimed that sharply delineated ('scharf abgegrenzt') was an "inherent" feature of terms.

Temmerman has since developed the sociocognitive approach to take in an ontological dimension (see Budin 2018, 94) and has, with her Brussels colleagues, undertaken important projects from an ontoterminological viewpoint, which will be commented on more fully in the section on Terminology and Cognitive Science. This takes the form of Termontology and Termontography (Temmerman 2007, 32),⁷ which aims at describing language within the context of human cognition,

7. "The termontography approach [...] is a multidisciplinary approach in which theories and methods for multilingual terminological analysis of sociocognitive theory are combined with methods and guidelines for ontology engineering. A clear distinction is made between conceptual modeling at a language-independent level and a language-specific analysis of units of understanding". (Temmerman 2007, 32)

building up from texts (thus a semasiological approach) and exploiting variation (Kerremans, de Baer and Temmerman 2010).

It is not surprising that the sociocognitivist approach has been interested in and critical of Wüster's conceptual analysis. Smith, Cuesters and Temmerman (2005, 649) point out inconsistencies in Wüster's definition of concepts and characteristics reflected in the ISO standards, which have been passed on in turn to SNOMED, the medical terminology database. As these authors point out, Wüster focused on technological terminology, that of human-made artefacts, not medical vocabulary: the aim of standardising terminology is part of the process of standardising manufactured objects, which is far from the aims of medicine. For medicine, they thus propose to scrap the concept as such and to work out a system to 'represent such universal types or kinds' (Smith, Cuesters and Temmerman 2005, 650).

4.3 Textual Terminology

J. Roche (2018, 56) posits a 'textual turn' in the history of terminology studies, a development which can be claimed to have profoundly changed research in the field. Indeed, the Textual Terminology approach and its rejection of many of the Wüsterian principles is in some way prototypical of the development of Terminology in general, since the status of the terminologists themselves had undergone a tectonic shift. In the days of Wüster and his predecessors, specialist lexicographers such as Schlomann (Schlomann-Lowe and Wright 2006), the terminologist was an expert, working with other experts. Bourigault and Slodzian (1999) and Condamines (2000, 16) take Wüster to task for not using a corpus, hardly a reasonable criticism since corpora as we know them now did not exist at the time. For terminographers of his era, everyday work was their corpus, including their own conceptualisation of the field they dealt with.

Everything changed when the terminologists were no longer experts but rather linguists (or *langagiers* as they are known in Quebec). The only way a linguist can access specialised knowledge is through texts, which may be oral but more particularly written. Fortunately, Corpus Linguistics was just coming into its own in the 1980s so that specialised corpora, initially difficult to access, could facilitate the extraction of knowledge and thus terms through texts. Chukwu (1996) was one of the first to realize that the linguist has no alternative than to build up cognitive structure from its textual base and Ingrid Meyer (Meyer, Bowker and Eck 1992) took this a step further by conceiving not just a term base but a terminological knowledge base fed by texts, Cogniterm.

As it happened, the case for Text-based Terminology was made cogently by a linguist clearly oriented towards the philosophy of science and epistemological

questions, Monique Slodzian (1995), who signed with Didier Bourigault the manifesto that became synonymous with the aims of the movement. Their condemnation of Wüster's theory was complete:

Il est illusoire de chercher à aménager la doctrine : le postulat d'une signification conçue comme discrète ou discrétisable, objectivante et permanente qui caractériserait le terme *a priori* est antinomique avec la terminologie textuelle. Les reformulations théoriques superficielles qui ont apparu ces dernières années sont vaines. [...] La terminologie doit sortir d'une sémiotique du signe fondée sur la triade terme/concept/référent qui la rend inapte à aborder le texte.

(Bourigault and Slodzian 1999, 31)⁸

One of the first realisations that handling large quantities of texts brought out was the omnipresence of variation in terminology. There was some confusion here about whether the General Theory of Terminology claimed that terms were intrinsically monosemic and unambiguous or whether they should be so. In addition, Wüster is accused of ignoring social needs and not recognizing that norms⁹ are in texts (Aito 2000; Condamines 2005, 2008a). Auger (1999) argues that corpora are needed for terminographical work in order to determine which possible variants are used by which socio-professional category and to guide recommendations.

Not only had the profile of the terminologist changed dramatically since Wüster's time, the aims of Terminology were also different. Much of the work of the 1990's was carried out within companies to keep track of in-house change and emerging trends, which necessitated a fine-grained interpretation of variation in the texts, only accessible through discourse analysis (Condamines 2005, 38, 2008a, 51). In addition, tracking change and emerging trends implied *ipso facto* the incorporation of the diachronic dimension into the method, a much shorter 'brachychrony' (Mair 1997, cited in Picton 2009, 17) that Wüster had rebelled against at a time when Linguistics was only just establishing synchrony as the default research setting, after language study had been etymology and Indo-European evolution.

8. There is no point in trying to reform the doctrine: positing a meaning conceived of as discrete or made discrete, objectivising and permanent, which characterises the term *a priori*, is completely at odds with Textual Terminology. Those superficial theoretical reformulations which have appeared over the last few years serve no purpose. [...] Terminology needs to abandon semiotics of the sign based on the triad term/concept/referent, which fails to equip it to account for texts (my translation).

9. It should be noted that *norme* in French represents both abstract language norms and concrete industrial standards.

4.4 Translation

The importance of Specialised Translation in the reception of Wüster's thought is generally less well covered than the orientations discussed above. Rogers (2015, 82, 95) regrets that the importance of terminology in Specialised Translation is all too often underestimated, which may go some way towards explaining this blind angle, especially as it was through translation that Wüster came to international attention. His relevance to translating in the German-speaking countries has never been in doubt however, as the numerous publications of Bühler (1982b, 2005) amply testify. She underlines the concomitance of the first paper on Terminology as a subject and the first on Translation Studies as such (Neubert 1972) at the third AILA conference in 1972. Wüster had been active in the International Federation of Translation or FIT on terminological questions since 1959 (Bühler 1982b, 426). On the theoretical plane, the parallels Bühler draws between Terminology and Translation as being both transdisciplinary "Referenzwissenschaften" is an original take on the relation between language and the world of specialised knowledge. Both require a conscious relationship with reference in order to assure efficient communication (Bühler 2005, 79).

This point is taken up by the terminologist Weissenhofer (1994) and the translation specialist Gerzymisch-Arbogast (1996), the latter expanding most usefully on concept structure at the textual level, as is only natural for a translator. In fact, this is a feature that surfaces in various unrelated studies, such as that of Béjoint (1988) quoted by Gledhill (2000, 24) positing term structure as revealed by texts.

In French-speaking countries Wüster's relevance to technical translation was generally recognised within his lifetime: Maillot (1969) credits Wüster as the reference for Terminology, but criticises his thesis as being more engineering than linguistic and in confusing polysemy and homonymy (Maillot 1969, 256).

Translation and more especially the teaching of specialized translation have had the effect of putting Terminology on the academic map at least in European countries, where most Terminology courses offered at universities are part of a translation syllabus (Humbley 2015). The realisation of the relevance of Terminology in all its forms to Specialised Translation has led to a 'rehabilitation' of Wüster's teachings, first in the analysis which Marc Van Campenhoudt made at the international conference on Wüster's legacy in 2006 and subsequently by others working in the same line, "a semic approach to the term paves the way to a terminographic practice which is very close to how Wüster defined it." Van Campenhoudt (2006, 6).

In 1997, Van Campenhoudt had minimized Wüster's place in the history of Terminology theory, invoking very reasonably the theoretical sophistication behind Captain Paasch's maritime dictionaries, though recognizing the relevance of Wüster's theory of the concept to the question of translation equivalence. This

point is developed in Van Campenhoudt (2006, 7), together with the remark that the closest embodiment of Wüsterian terminography takes the form of the students' end-of-year glossary, kingpin in the training cycle.

The necessity to adopt an onomasiological approach to specialised translation is also underlined by Héba Medhat-Lecocq in a series of studies carried out in wide-ranging domains: in Architecture (Medhat-Lecocq 2014) and in Legal Translation (Medhat-Lecocq 2019), where an analysis of characters is an essential step in finding equivalents or dealing with their absence. Frérot (2018) speaks specifically of the rehabilitation of Wüster's principles in terminography, drawing on her experience for translating for WIPO (World Intellectual Property Organisation) and her participation in launching its terminology portal. She pleads for the relevance of a component of standardization necessary to assure clear communication in an institutional setting, a requirement noted somewhat earlier... but in a Danish context, by Lervad (2009, 74–75). From the evidence marshalled above, it would seem that Specialised Translation is perhaps the field in which Wüster's teaching has been most systematically acknowledged.

5. Issues in Terminology research: The primacy of the concept

Different directions in terminology research have thus defined themselves in relation to the General Theory of Terminology, which, even when it was no longer 'dominant', still served as a yardstick to measure how the field was developing. This was often done as a negative reaction or a protest about perceived terminology 'dogma' (cf. 'centralized control from Vienna', Cabré forthcoming, 34), although the developments could equally be viewed as carrying on the traditions rather than overthrowing them. There are several items of Terminology theory that have proved contentious but one major theme can perhaps adequately represent Wüster's thought and how it was criticized: the primacy of the concept.

This point is central to Wüster's theory, as Myking (2008, 51) makes clear.¹⁰ The logical step from the primacy of the concept is the onomasiological approach, which was doubly natural in the mid-twentieth century: firstly, because as a specialist, that is how a field is apprehended. Experts model their knowledge by comparing with similar concepts and by identifying the characters or features that distinguish them. Secondly, in the course of their work, in particular for standardization, this requires crafting a definition, indispensable for term banks but often left out of descriptive terminology.

10. Jede Terminologearbeit geht von den Begriffen aus. All terminology work starts from concepts. Wüster ([1979] 1991, 1)

For a textually-based terminology, as has been claimed, the starting point must be terms extracted from texts, therefore by definition a semasiological approach. Indeed, much of the most fruitful research carried out around the turn of the millennium was focused on using this method to automatize as far as possible extracting terms from texts. But this does not preclude a second, onomasiologically oriented, processing of terms thus collected, typically presented in the form of graphic representations of knowledge structure (see Nuopponen 1994a, 2018b).

It will be remembered that Wüster latterly posited Terminology as an interdisciplinary science including among other fields what he called Ontology. The definition of *ontology* has of course evolved radically since the 1970s. C. Roche gives a revised definition and calls its terminological counterpart *ontoterminology* (C. Roche 2007, 2012a).¹¹

Among the linguists, semantician Pierre Lerat recognized early the relevance of the ontological dimension of Wüster's thought:

[...] Wüster a mis l'accent sur une autre intersection de la terminologie et de la philosophie : l'ontologie. Par exemple, sont des relations ontologiques les dépendances conceptuelles d'ordre spatial ou temporel, ou encore la relation “*partie de*”. [Lerat renvoie explicitement au manuel de Wüster]. [...] Quand les inférences sont nécessaires ou très probables, elles se prêtent au calcul, comme le montrent les systèmes experts. (Lerat 1988, 15)¹²

The interest in the concept as the starting point of terminological analysis however came as much from linguists, notably those who created the Terminologie et Intelligence Artificielle (TIA) group around Anne Condamines, as from specialists in Artificial Intelligence, trained in Linguistics and working with linguists. First and foremost are Christophe Roche and more recently Julien Roche.

11. There is also a link between standardization: C. Roche was for years the project leader of the ISO 1087 standard on “Terminology work and Terminology science – Vocabulary”, one of the core standards of the TC 37, and terminology in the use of analysis by characters. Wüster's (1974) conception of ontology was worked out in relation to concept systems and in contradistinction to logical concept systems but as from the 1980s, ontology was developed so that modern terminological development could not exist without it (Budin 2018, 94–95).

12. [...] Wüster stressed another intersection between terminology and philosophy, namely ontology. Some examples of ontological relations are those of spatial or temporal conceptual dependency or the ‘part of’ relationship [Lerat refers explicitly to Wüster's textbook. [...] When inferences are necessary or highly probable, they can be used to calculate, as expert systems show (my translation).

Il ne peut y avoir de terminologie sans connaissances. La construction d'un système notionnel est cependant ardue. Elle l'est encore plus qu'elle doit expliciter des connaissances souvent tacites qui, rarement décrites dans les documents scientifiques et techniques, rendent la présence des experts et leur collaboration au travail terminologique indispensables. (C. Roche 2008, 68)¹³

C. Roche (2012a, 2627) draws attention to the complementarity nature of Terminology and Artificial Intelligence through definitions written in natural language in order to clarify or standardize concepts. He claims that Terminology is relevant to Artificial Intelligence by its striving to represent concepts in natural or semi-formal language, as against AI, which uses formal language in order to manipulate or compute concepts. In other words, this feature of Wüster's theory that Socioterminology rejects is precisely what is relevant to a modest contribution to thought on concept representation in AI. Roche proposes an *ontoterminology* i.e., "a terminology whose conceptual system is a formal ontology relying on epistemological principles" (C. Roche 2013, 2628). In this, ontoterminology is different from but complementary to termonography (Temmerman 2000), the latter focusing on specialised vocabulary and giving priority to the semasiological approach. As far as terminography is concerned, there is a case to be made for dialectical application of both (Myking 2005, 179; Laurén, Myking and Picht 1998, 218).

This [reappraisal of Terminology principles] is not intended as yet another criticism of traditional terminology [...] On the contrary, our contribution is distinctly Wüsterian in its scientific approach. [...] current thinking and practices in the field of terminology must be integrated into a reviewed version of ISO 704 standard, in particular to take the notion of ontology for terminology into consideration [...]. (C. Roche 2012b, 17–18)

C. Roche (2012b, 20) stresses the relevance of identifying characteristics as a basis for concept analysis. J. Roche (2018, 15) in his thesis, declares a Wüsterian filiation and places the first part of his work in the framework of the General Theory of Terminology. The point of contact between Terminology and Cognitive Science is found in Ontology.

Wüster's approach to concept representation naturally struck a chord with specialists of the cognitive sciences, since the importance of figurative and therefore non-linguistic representation was demonstrated, for example in the 1968 dictionary:

13. There can be no terminology without knowledge. But constructing a conceptual system is a hard task. It is even harder when it comes to making explicit what is often tacit knowledge, seldom consigned to scientific and technical documents, which renders the collaboration of experts indispensable in terminology work (my translation).

Comment ne pas citer l'exemple du dictionnaire multilingue de la machine-outil de E. Wüster, fondateur de la terminologie moderne, où les connaissances du domaine sont représentées sous la forme de dessins techniques exprimés dans un langage figuratif et normalisé. (C. Roche 2013, 23)¹⁴

C. Roche (2013, 24) quite logically, traces this non-verbal representation of knowledge back to Diderot's *Encyclopédie*, where no less than eleven volumes were given over to graphic representation.

Concept modelling has been a long-standing speciality of some Scandinavian terminologists. As previously mentioned, Nuopponen (1994a, 2018b) is perhaps the best known. They explicitly claim descendancy from Wüster, explaining that their aim is to continue, not overthrow, his analysis, for example that of causality for Nuopponen (1994a, 532). The cause-and-effect relationships are touched on in Wüster's writings but more fully developed in Nuopponen's. This orientation has been characterized by Myking (2001a) as loyal and radical, underlying the continuity in ambition from Wüster's teaching to the present.

6. Misapprehensions: *Ist-Norm/Soll-Norm*

Misapprehensions concerning Wüster's thought have been analysed by several scholars. Foremost among these are questions of synonymy, the metaphor, the analysis of terms in context and more generally the importance of description in term analysis. The criticisms made have themselves been criticized for being only partly founded (Antia 2001), and taken out of context, since they give a distorted idea of the original thrust. One feature of Wüster's thoughts about standardization that is seldom taken into account by his critics is the distinction proposed between what he called the *Ist-Norm*, or the standard as it is, and the *Soll-Norm*, or standard as it should be. This may be what inspired Cabré (Forthcoming, 9) to call it 'a terminology of desire' rather than 'a terminology of reality', though the implications are rather different.¹⁵ The 'standard as it represents the language forms actually used in the place of work and which need to be taken into account in terminology work'.

In many cases it is a term that has simply evolved (*gewachsen*, *grown* is the image Wüster uses [(1959–1960) 2001, 316]) and which may need to be "constituted"

14. How can we fail to call to mind the example of the *Multilingual Machine-Tool Dictionary*, by E. Wüster, the founder of modern terminology, where the domain knowledge is represented by industrial drafts expressed as a figurative, standardized language? (my translation).

15. Cabré (forthcoming) 'Wüster': ... a theory which defended that terms were biunivocal and did not present variation.

or “constructed” (Samain 2008) so that it can be properly motivated as an official designation. There is no implication in Wüster’s writing that motivated, unambiguous terms are the usual form in which terms appear in specialized discourse, written or oral. Terminologists who seek to discredit Wüster by presenting examples of ambiguous or confusingly motivated terms which may be a *de facto* norm in specialised circles are in fact missing the point.

7. Conclusion

The very nature of research articles is to present innovation in contradistinction to the state of the art. It was thus only natural that Wüster’s work should have been chosen to represent Terminology as it was first conceived as a discipline and against which innovation could be measured. Wüster’s reputation, gained as has been shown by his institutional achievements, was such that even researchers who were only acquainted with the few articles available in languages other than German were keen to present their work in relation to what was seen as accepted practice. This in turn gave rise to the supposition that terminology research was *pro* or *anti*-Wüsterian. Adopting this narrow viewpoint though can lead to missing the essential contribution that Wüster made. This contribution can be best appreciated by adopting a broad-based approach integrating Terminology’s profoundly interdisciplinary nature on one hand and a non-judgmental attitude on the other, going beyond acceptance or rejection to embrace all reactions to his works.

It has been argued here that Eugen Wüster’s heritage is to be found as much in what he created as what he wrote. His role in founding the methodology and the institutions of standardizing terminology, embodied in the work of ISO/TC 37, is fundamental from this point of view. What is more, his initiative in founding entities devoted to various aspects of terminology work was important in institutionalising both the discipline and the associated activity (links with Documentation, Specialised Translation, Technical Lexicography, Cognitive Science etc.). Leading on from this is the interdisciplinary nature of his work, which corresponds to the aims of Applied Linguistics but which can lead to conflicts in academic loyalty, in particular to Linguistics: is Terminology part of Linguistics or is it a broader, overlapping field?

It has also been argued in this review of the reception of Wüster’s thinking that scholars have reacted principally to issues of Linguistics, first and foremost the relationship between language and reality, in particular – a major innovation – of how language participates in scientific and especially technical thought. This in turn entails discussion of several basic linguistic concepts: polysemy, synonymy, not to

mention aspects often neglected by linguists such as acronymy. All these analyses were carried out in the perspective of solving problems of communication, and in this respect Wüster fully lived up to the ideals of Applied Linguistics in the broadest and noblest sense of the term.

Acknowledgements

The author would like to offer heart-felt thanks to the following scholars, who helped greatly in preparing this chapter: Pierre Auger, Danielle Candel, Valérie Delavigne, Susanne Lervad, Mojca Pecman, Heribert Picht, Christophe Roche. The views expressed, however, are strictly those of the author.

General principles of Wüster's General Theory of Terminology

Danielle Candel
Université de Paris

This chapter introduces Terminology as presented by Wüster, based on his writing as well as on unpublished notes and letters. First an engineer specialized in sawing and woodworking tools, Wüster became interested in language, and lectured at the University of Vienna in Lexicology, Lexicography, Terminology and Standardization. His revolutionary ideas on Terminology triggered reactions which led him to explain his views in greater depth. This chapter discusses concepts that arose as he developed his theory. As reflected in his comments on phraseology, standardization might not be incompatible with “descriptive terminology”: Even though he advocated the *Soll-Norm* (norm as it should be) in Theoretical Linguistics, he accepted the *Ist-Norm* (norm as it is) in Descriptive Linguistics.

Keywords: biunivocity, concept, denomination, descriptive, prescriptive, Ist-Norm, Soll-Norm, standardization, Terminology, theory

1. Introduction

Analyzing the field of Terminology implies studying Eugen Wüster's (1898–1977) pioneering role and activity in this discipline, which is not always as well-known,¹ or as well understood² as it should be. This chapter presents aspects of Wüster's pioneering and important work with a view to exploring what is called *terminology* in his writings. This is based on the way that Wüster expressed himself but also on his theories when he discussed his choice of terms. In addition to Wüster's remarkable

1. An extensive list of Wüster's work is given in Oeser and Galinski (1998).

2. See also Humbley (2000, 2004, 2018b, chapter in this volume), which also helps to fill this gap and offers enlightening insights on Wüster's role in terminology.

number of articles and books, it is also instructive to examine his legacy³ including some numerous and valuable unpublished work. The Wüster Archive at the University of Vienna gives an idea of the broad spectrum of his research activities (Bühler 1982a). It covers topics such as spelling reform, standardization, library science and documentation, decimal classification, Terminology and Lexicography, dictionaries for special purposes, Applied Linguistics, and Esperanto – the field that initially aroused his interest in language.

Although Wüster was not exclusively interested in Terminology, he highlighted its specificity compared to other subjects that he was working on. In his ground-breaking work, he felt that it was important to provide a general set of terminological principles. In his dissertation, Wüster ([1931], 1970) proposed the scientific basis for his *General Theory of Terminology* (GTT), as well as for the terminological work of the International Organization for Standardization (ISO). For this reason, he is known as the pioneer of international standardization efforts. Concept formation was evidently as important to Wüster as word formation, and Logic was as important as Linguistics. It is thus interesting to examine both the prescriptive and descriptive aspects of his work. We would also like to underline the importance of the concepts of prescription and description in Terminology and Lexicography. Whereas lexicography is generally more descriptive, terminology has been traditionally regarded as being more prescriptive.

International research in Terminology began in Germany as early as 1931 with the *Verein deutscher Ingenieure*, the Association of German Engineers, who held a meeting with researchers in Linguistics (Wüster [1974] 1993). The International Association for Applied Linguistics (*Association internationale de linguistique appliquée*, or AILA), created in France in 1964, promoted the development of Terminology. Although some people were convinced that Terminology (*Terminologielehre* in German) would simply become a part of Linguistics, others felt that Linguistics mainly dealt with general language, and thus was less interested in specialized concepts. Their assertion was that technical languages are artificial structures. The founding of Infoterm by the UNESCO in Vienna in 1971 largely helped to disseminate Wüster's work. During this period, Wüster corresponded with numerous authors, linguists, terminologists and publishers, and his exchanges are a rich testimony of his pioneering work.

The rest of this chapter is organized as follows. Section 2 describes intellectual life in Vienna previous to and during Wüster's time. Section 3 discusses how he began to work in Terminology. Section 4 then focuses on stability in Terminology,

3. Thanks to Gerhard Budin, we were able to work at the Wüster Archive in Vienna in December 2003 (at that time at Lammgasse). See also Budin (1998) and Nedobity (1998).

and Section 5 deals with terminological variation. Finally, Sections 6 and 7 analyze Wüster's position and how this position is now interpreted.

2. Intellectual life in Vienna before and during Wüster's time – towards the Vienna School of Terminology

It is instructive to look back at the sources of languages for special purposes and, more specifically, of modern Terminology as a new discipline and a language policy tool in Western Europe (particularly in France and Austria), in the period between the two world wars and immediately after 1945 (see Humbley 2000). This underlines their role in society and situates the emergence of Terminology in its social and historical context (Nedobity 1984; Budin 2007).

Here it is necessary to mention the influence of the Vienna Circle (*Wiener Kreis*). The Philosophical Ontology Movement (1923–1936) characterized by Positivism and the Philosophy of Science, is a cross-disciplinary field combining Phenomenology, Cognitive Psychology, Logic, Linguistics, and Epistemology. It was the origin of Wüster's new discipline, Terminology.

Sources include the works of Kant and Mauthner. Mauthner was influenced by Mach, for whom philosophical issues were regarded as language problems because scientific language is inadequate. Wüster also read Locke and Schopenhauer. Mauthner (see Nedobity 1984 and Budin 2007) criticizes the idea that language denotes concepts, as names are only metaphors of human sensory impressions. Contrary to Mauthner, and in line with Wittgenstein, Wüster claimed that language, both in Natural Sciences and in Philosophy, gives access to reality. Nedobity (1984) underlines the influence of Mathematics, Physics or Mechanics on language: Mach, Boltzmann, and the mathematician, Hahn, influenced Schlick, the founder of the Vienna Circle, while Carnap joined the group in 1926. Their discussions had a marked influence on Wüster.

Wittgenstein believed in a descriptive language in Physics (*eine darstellende Sprache*, see Nedobity 1984, 44), whereas Mauthner did not. Representations and images, as mathematical models, were considered important for language in theoretical Physics according to Hertz as well as Boltzmann. Wittgenstein thought that this was applicable to general language (images from facts), and Wüster, to language for special purposes, which derives concepts from objects. For Wüster there was a clear delimitation between semantics and logic and on this matter he differed from Wittgenstein. He borrowed the language-critical and philosophical approaches of the representatives of the Vienna Circle and combined them with the findings of prestigious linguists in order to create a practice-oriented science.

This led him to focus on specific domains and initiate work on standardization. At the end of the 18th century, naming rules became necessary for domains such as Botany, Zoology, Chemistry, and Medicine because this was the only way to increase the accuracy of linguistic designations. And standardization was the quickest way to reach this goal. Multidisciplinary and multilingual terminological principles were thus needed for science, technology, and economy. These ideas that were initially developed in Central Europe were also disseminated in the Soviet Union (Felber 1998) in the form of a Russian translation, by Drezen, a researcher close to Wüster.

The International Standards Association (ISA) was created in 1926⁴ and the ISA 37 committee, *Terminology* was set up in 1936. ISA became ISO (International Organization for Standardization) after World War II, and was replaced by ISO/TC 37 *Language and Terminology* initiated by Wüster in 1951. Wüster's work on the international standardization of technical language (1931) helped to establish the Technical Committee 37 of ISO, whose task was to make recommendations for Terminology. Wüster's contribution, *Internationale Sprachnormung*, is regarded as the standard work of Terminology, also used for DNA (initially DIN), the German section of ISO.

Wüster thus created the prerequisites for terminological principles and standardization. In 1955, he began to work on naming as a way to enhance international standards such as ISO 1087, *Vocabulary of Terminology* and ISO 1990. He then wrote the dictionary *Grundbegriffe der Terminologielehre*, considered to be the continuation of his *Internationale Sprachnormung*.

Wüster regarded the objective world as individual objects to which concepts were assigned and given names (Nedobity 1984). Wüster's classification theory, a basic component of his General Terminology, has four fields: subject area, language, language access, and language overview. He presented his Terminology as a new interdisciplinary field, at the 3rd AILA Congress in 1972. He was concerned with the Philosophy of Language, particularly with the relationship between concept⁵ and object. Weisgerber's influence is evident in Wüster's title, *Worten der Welt* [Wording the World], which refers to the process of creating relationships between perceptible objects and concepts on the one hand, and meaning and signs on the other: *Das Zeichen ist somit ein perzipierbarer Ersatzgegenstand für den Begriff*⁶ (Oeser and Budin 1999, 2173).

In *Das Worten der Welt* ([1959–1960] 2001), he explains the links that arise when wording the objective world (see also Sager 2003). *Wording* is characterized

-
4. AFNOR (France) was created the same year.
 5. We use *concept* instead of *notion* (as Wüster himself does when he writes in French).
 6. The sign is thus a perceptible substitute for the concept (our translation).

by relations between individual objects and concepts, and between signs and meaning. An *individual object* is a physical or mental object that is perceived. Its representation remains, even when it is no longer present. If it is not visual, or if it is just a memory, it is a concept, which might be an individual concept. Individual objects may be confused with each other and be given the same name. Individual concepts of these objects may be blended and their differentiating characteristics disappear. Different individual concepts may be instances of a general concept. Since concepts are mental objects, existing only in people's minds, they are immaterial.

What Wüster calls *material and immaterial concepts* designate concepts of material and of immaterial [individual] objects. He also explains what he calls *systems of individual concepts* (*Bestandsysteme*), *sign and meaning* (*Zeichen und Bedeutung*) (Regulas' *Sinnform* and Martis' *innere Form*). These are a few examples of Wüster's definitions of certain notions that contributed to the creation of Terminology as a science.

According to Wüster ([1974] 1993), three aspects were very important. The first is that terminologists deal with concepts, whereas linguists are more interested in word content. The second is that vocabulary is more important than grammar. The third is that synchrony is the most important dimension in Terminology. The importance of Wüster's contribution was recognized as early as 1931. By explaining his view of Terminology, he effectively created what would later be called the Vienna School of Terminology.

3. How Wüster situates himself

In his writings, Wüster, driven by a desire to be more explicit, sometimes seems to be defending different views or introducing nuances, an aspect that deserves further examination.

Wüster points out that the conceptual system was initiated in Linguistics by Saussure, and Weisgerber and Trier, founders of Structural Lexicology. It was put into practice in Wartburg and Hallig's *Begriffssystem* (Wüster [1974] 1993). He also comments (Letter to Wartburg, 30/11/1952, typescript) that one should not only be interested in comparing the *Begriffssysteme* of different national languages, but also in comparing those of different user groups within each language. These groups can thus be analyzed, based on general language, languages for special purposes, special language layers etc. Wartburg (letter to Wüster, 19/12/1952, typescript) concludes that what matters to Wüster is the recording of reality, whereas he is more interested in how the factual is reflected in the language. Nevertheless, Wüster is also concerned about linguistic variation in specialized language.

Wüster explains what ‘terminology’ means to him. Terminology work has three levels: (i) research (exploration of the *Ist-Norm*); (ii) layout (norming); and (iii) presentation (as in dictionaries) (Wüster 1975, 34). He states that for the leader of the German Federal Language Office, only the second level, the *Soll-Norm*, counts. For Dubuc (1975), who was in charge of the Terminology Bank at the University of Montreal, standardization, though important, is closer to Lexicography than Terminology. Wüster, however, positions himself between these two extremes:

Was ist unter “Terminologie” zu verstehen? Terminologie als Ausgangsbegriff unserer Überlegungen ist das Begriffs- und Benennungssystem eines Fachgebietes; also der Gegenstand fachsprachlicher Lexikologie. (Wüster 1975, 33)⁷

Wüster feels that his title, *Allgemeine Terminologielehre*, is quite ambitious. In his view, *Terminology* sounds too general since it really has three meanings. First, *terminology* refers to the concept and meaning system of any area of expertise. Secondly, it is the material representation of this terminology, just as systematic dictionaries are often called *terminologies*. Thirdly, it also refers to the discipline dealing with the specialized lexicology of a domain, also known as *Special Terminology*. At this point, he notes that *Terminologielehre* presents a tautology since *-logie* and *Lehre* have similar meanings:

Wer unbedingt die Zusammensetzung ‘Terminologielehre’ vermeiden will, könnte etwa sagen: ‘Theorie der Terminologie’. Aber das engt den Begriff etwas ein. (Wüster [1974] 1993, 332)⁸

This conclusion says a lot about how Wüster’s message should be understood. His *Terminologielehre* does not absolutely refer to a theory, nor is the *allgemeine Terminologielehre* a ‘General Theory’. Instead *die Terminologielehre* designates the science, namely, the discipline of Terminology.

As mentioned above, the third meaning of *terminology* is the set of compliances resulting from specific terminologies belonging to many specialist domains and languages. The *Allgemeine Terminologielehre* is not a set of special terminologies. Its relation to special terminologies is similar to that of General Linguistics to the linguistics of individual languages. Finally, as Wüster observes, *Allgemeine Terminologielehre* could be replaced by *Terminologische Grundsatzlehre* (Basic Principles of Terminology). For Wüster, these three meanings of *terminology* are absolutely necessary.

7. What does ‘terminology’ mean? Terminology as the starting point of our considerations is the concept and naming system in a domain; and thus the object of technical lexicology (our translation).

8. If you absolutely want to avoid the term ‘Terminologielehre’, you may say: ‘theory of terminology’. But this narrows the term somewhat (Wüster [1974] 1993, 332) (our translation).

Concerning knowledge in Terminology (*Terminologiekennntnis*), there is either a simple terminology science (*einfache Terminologiekunde*), covering the practical, non-systematic knowledge of a terminology, or a terminology theory (*Terminologielehre*), which specifically refers to the systematic overview of one or several terminologies, and constitutes the knowledge of the naming principles which make up the system (*die Kenntnis der systembildenden Benennungsgrundsätze*) (Wüster 1975, 35).

According to Wüster, terminology that only concerns one domain or one language is known as *spezielle Terminologielehre* (special terminology):

Durch Abstraktion aus mehreren oder vielen speziellen Terminologielehren entsteht eine Allgemeine Terminologielehre, d.h. eine überfachliche oder übersprachliche. (Wüster 1975, 35)⁹

Wüster ([1974] 1993) highlights a major difference between formal sciences such as Linguistics, Logic, Ontology, and Computer Science, and sciences such as Biology, Physics, and Electrotechnology:

Die Allgemeine Terminologielehre ist für den lexikalischen Teil der Fachsprachen das, was die Allgemeine Sprachwissenschaft für die Allgemeinheit ist. (Wüster 1975, 35)¹⁰

Felber ([1986], 1993) alludes to Wüster's ([1974] 1993) own terminology ("Die Terminologiewissenschaft – er nennt sie Allgemeine Terminologielehre")¹¹ to remind us that the ATL is just the science of Terminology. "Terminology research based on the physical sciences is best represented by Wüster", who makes a distinction between general terminology and special terminologies. General Terminology is where Linguistics, Logic, Ontology, Computer Science and Physical Sciences overlap. Special terminology studies deal with the regularities of terminology in specific fields or languages.

Dealing with terminological lexicography, Wüster highlights the relations between concepts in any technical dictionary (*Fachwörterbuch*). As for terminologists, their terminological knowledge is underlined in at least one specialist field. Regarding international standardization in terminology for theoretical linguists, general terminology is required. For people who are not terminologists, simple terminology science is sufficient. However, mastery of a subject field requires knowledge of the terminology in that field (*einzelfachliche Terminologielehre*) (Wüster 1975, 35).

9. By abstracting from several or many special terminologies, a general terminology is created, i.e. a supra-disciplinary or supra-linguistic one (our translation).

10. The General Terminology is for the lexical part of technical languages what General Linguistics is for general language (our translation).

11. Terminology science, which he calls *General Terminology Studies*.

4. Stability in Terminology

Terminologists should analyze signs and their relations to concepts and conceptual systems, and account for regularities in systems of concepts, as a part of logic. Wüster's aim was to provide naming principles and rules (terminology being mostly restricted to vocabulary and to synchrony), and thus he proposes a set of far-reaching terminological definitions.

As claimed by Kocourek (1991), Wüster's classification of relationships between concepts within a terminology (*Zusammenhang zwischen Begriffen*), which is a model for a number of terminological standards, is the best known: subordination/superordination, coordination, logical scale, ontological scale, logical series, and ontological series. Terminology examines all concepts of a subject area in context, i.e., as part of a conceptual system:

Jede Terminologearbeit geht von den Begriffen aus (...). Das Reich der Begriffe wird in der Terminologie als unabhängig vom Reich der Benennungen (= Termini) angesehen. ([1979] 1985, 1)¹²

Terminologists refer to concepts whereas linguists refer to word meanings. For terminologists, only denotation is important as opposed to connotation (Wüster [1979] 1985, 2). A concept is derived from various objects. It is a type of mental classification, and thus is a means of communicating. The concept is an element of thought.

Wüster mentions the following types of conceptual relation: (i) logical subordination (*apple/fruit*); (ii) logical coordination (*apple/pear*); and (iii) diagonal relation (*mammal/snail*) ([1979] 1985, 9–10). Concepts are related vertically or horizontally (see chapter by Nuopponen in this volume). Wüster distinguishes three kinds of logical subordination and explains determination, conceptual conjunction and conceptual disjunction. According to Sager (2003, 275), an example of *determination* is “the concept ‘plant’ restricted to the concept ‘tree’”. *Conjunction* occurs when the intensions of a concept are united. As for *disjunction*, in regard to ‘man’ and ‘woman’, it furnishes the superordinate concept ‘bisexual human being’.

Specialized dictionaries begin with generic concepts and proceed to more specific ones. The most important relations deal with the whole and its parts (meronymy), or with the parts themselves, with inclusive subordination or the coordination of parts.

Wüster explains structural relations between three or more concepts, between vertical and horizontal relations, and between combinations of parts. Inherent or

12. All terminology work is based on concepts (...). The domain of concepts is considered in Terminology as independent from the domain of denominations (= terms) (our translation).

relational classification features help to categorize concepts. There are equivalent or ontologically equivalent features. Wüster explains how to prioritize them while classifying them, depending on the type of feature, such as the level of equivalence, or loss of equivalence. He analyzes specific concept systems and their representation through conceptual diagrams (Wüster [1979] 1985, 5–19).

In practical terminology work, Wüster made an early distinction between concept classification and subject classification. He emphasized the need to prioritize the logical-philosophical foundations of Terminology over general linguistic ones by considering the realm of concepts as independent of the realm of denominations (Oeser and Galinski 1998, 8–10).

In Terminology, one first considers the concept and the relations between concepts and between individuals. There are logical conceptual systems, relations, comparative relationships, and linking of terms. However, in addition to relations between concepts, there are connections to denominations and general word formation. There are also logical linkings of concepts to parts of word groups, such as the following:

gekennzeichnet durch für Determination
und (zugleich) für Begriffs-Konjunktion
oder für Begriffs-Disjunktion. (Wüster [1974] 1993, 359)¹³

As far as the definitions of *terminology* are concerned, and as previously mentioned in Section 3, Wüster underlines the polyvalency of *terminology* (Wüster [1974] 1993, 333), first, as a specialized vocabulary:

Erstens bedeutet ‘Terminologie’ das Begriffs- und Benennungssystem irgendeines Fachgebietes; also, grob gesagt, einen **Wortschatz** mit den zugeordneten Bedeutungen. Hieran schließt sich eine Nebenbedeutung, die nur anmerkungsweise erwähnt werden soll. Und zwar wird auch die geordnete **Darstellung** einer solchen Terminologie, z. B. ein systematisches Fachwörterbuch, oft ‘Terminologie’ genannt. Genau so wie die internationalen Wörter ‘Vokabular’ und ‘Lexikon’ in ihrer französischen und englischen, ja sogar deutschen Ausprägung gleichzeitig einen Wortschatz und bestimmte Arten von Wörterbüchern bedeuten.¹⁴

13. *Identified by* for determination, *and (simultaneously)* for concept conjunction, *or* for concept disjunction (our translation).

14. First, ‘terminology’ represents the “Begriffs- und Benennungssystem” of any domain; roughly speaking, a **vocabulary** with the meanings assigned. This is followed by a secondary meaning that should just be mentioned. The ordered **presentation** of such a terminology, e.g. a systematic specialized dictionary, is often called “terminology”. Just like the international words ‘vocabulary’ and ‘lexicon’ in their French, English, and even German forms mean a vocabulary and certain types of dictionaries at the same time (our translation).

He then defines *Terminology* as a discipline (*Terminologielehre*), for example in the *Brockhaus Enzyklopädie* (1973), as the science of terminology, fundamental terminological principles, and specialized lexicology. A special terminology relates to a single subject area, such as Chemistry or Economics. General Terminology includes basic facts and principles, which are common to all or to many terminologies and languages. It encompasses regularities abstracted from special terminologies for many subject areas and languages. Drozd (1971) comments:

die Terminologielehre ist (...) ein Fach, dessen Stellung innerhalb des Systems der heutigen Wissenschaften nicht ganz eindeutig ist. (...) Viele Wissenschaftler sind bestrebt, diese Frage der Stellung der T-lehre zu beantworten. Meist wird die Problematik der T-Lehre von der angewandten aber auch von der speziellen Sprachwissenschaft mitbehandelt. Es kann sich um ein Teilgebiet der Lexikologie, der Syntax oder der Semantik handeln. Die "Sprache der Wissenschaft" ist schliesslich auch Gegenstand der Logik, der Sprachphilosophie oder der Kommunikationsforschung. (Drozd 1971, 2)¹⁵

Die zweite Hauptbedeutung von 'Terminologie' ist die **Lehre** von der Terminologie eines bestimmten Fachgebietes in einer bestimmten Sprache; das ist soviel wie die fachsprachliche Lexikologie dieses Fachgebietes. Seit einer Anzahl von Jahren sagt man hierfür auch 'Terminologielehre', genauer: 'spezielle Terminologielehre'. (...) Wer unbedingt die Zusammensetzung 'Terminologielehre' vermeiden will, könnte etwa sagen: 'Theorie der Terminologie'. Aber das engt den Begriff etwas ein.

(Wüster [1974] 1993, 332)¹⁶

And he defines *Terminology* as basic terminological principles. According to the DIN Standard 2342, 1986, *allgemeine Terminologielehre* means:

fachübergreifende oder sprachübergreifende Terminologielehre, die durch Abstraktion aus mehreren oder vielen speziellen Terminologielehren entsteht.

(Arntz and Picht 1989a, 6)¹⁷

15. Terminology (...) is a domain whose position within the system of contemporary sciences is not entirely clear. (...) Many scientists try to answer this question of the position of terminology. In most cases, the problems of T-science are dealt with in Applied Linguistics but also in Special Linguistics. It can be a sub-area of Lexicology, Syntax or Semantics. Finally, "language of science" is also the subject of Logic, the Philosophy of Language or Communication research (our translation).

16. The second main meaning of 'terminology' is the **study** of the terminology of a particular field in a particular language, similar to the technical lexicology of this field. For a number of years this has also been called 'terminology', more precisely: 'special terminology'. (...) Whoever absolutely wants to avoid the combination 'Terminology' could say, for example: 'Theory of Terminology'. But that narrows the concept a bit (our translation).

17. multidisciplinary or cross-lingual terminology, which results from the abstraction from several or many specific terminologies (our translation).

Each specific terminology results from one terminology in one domain in one language:

Eine dritte Bedeutung des Wortes 'Terminologie' ergibt sich, wenn man aus den speziellen Terminologielehren für viele Fachgebiete und Sprachen gemeinsame Gesetzmäßigkeiten abstrahiert. Das Ergebnis ist die 'Allgemeine Terminologielehre'. Darunter ist also nicht etwas eine Summe von speziellen Terminologielehren zu verstehen. (...) Statt 'Allgemeine Terminologielehre' kann man auch sagen 'terminologische Grundsatzlehre'. (Wüster [1974] 1993, 332)¹⁸

However, Wüster is particularly well known for stating how things ought to be in terminology (*Soll-Norm*). Among his most characteristic statements, the prescriptive *Soll-Norm* is how things should be, as opposed to the descriptive and spontaneous *Ist-Norm*, which is how things are (see also Morgenroth (1990, 79), Wüster ([1974] 1993, 337)):

Soll-Norm:

In der Terminologie (...), mit ihrer ungeheuren Fruchtbarkeit an Begriffen und Benennungen, führt die freie Sprachentwicklung zu einem untragbaren Durcheinander. Darum haben die Terminologen schon am Anfang unseres Jahrhunderts – auf einigen wenigen Fachgebieten sogar schon früher – begonnen einheitliche Begriffe und Benennungen zu vereinbaren. Sie nennen eine solche Vereinbarung ebenfalls 'Norm'. (...) man [hat] eine solche vereinbarte Norm als 'Soll-Norm' bezeichnet. In der fachlichen Hochsprache werden Soll-Normen ziemlich schnell zu Ist-Normen. (Wüster [1979] 1985, 2)¹⁹

The *Soll/Ist* opposition works for definitions as well. Wüster distinguishes general, intensional, and extensional definitions as well as descriptive and prescriptive ones: *Ist-* and *Solldefinitionen*. The first asserts what is, as reflected in general language, and the second, what should be, as in Mathematics or in standardization (Wüster [1979] 1985, 29 and 31). Opting for a strict attitude, Wüster regrets, for instance, that the conceptual and expressive sides are not sufficiently differentiated in Linguistics:

18. A third meaning of the word 'terminology' arises if, from the special terminologies, one abstracts common regularities for many subject areas and languages. The result is 'Terminology'. So this is not something to be understood as a sum of special terminologies. (...) Instead of 'general terminology', one may also say 'basic terminology' (our translation).

19. In Terminology, with this immense fertility in terms and denominations, the free development of language leads to unbearable confusion. This is why terminologists as from the beginning of the twentieth century – and in some domains even earlier – reached an agreement on homogenizing concepts and namings. Such an agreement is also called "norm". (...) such an agreed norm has been called a *Soll-Norm*. In formal technical language (*Hochsprache*), Soll-Norms become fairly quickly Ist-Norms (our translation).

die unzureichende Trennung der begrifflichen von der Ausdrucksseite.

(Wüster [1974] 1993, 365)²⁰

Underlining stability in Terminology, Wüster uses the term *biunivocity* (*Eindeutigkeit*), which is fundamental to Mathematics and Formal Logic. This term means reversible unambiguity or ambiguity in both directions. In other words, the relation between a term and its designation is unambiguous if this term only has one designation and if this designation has no other meaning (Gerzymisch-Arbogast 1996, 10). Around the beginning of the 20th century, terminologists began to unify concepts and designations. This is also called *Norm*, the “Soll-Norm” or ‘prescriptive norm’:

[Es] ist in der Terminologie zu verlangen, dass sie, um einen Ausdruck aus der Mathematik zu gebrauchen, eineindeutig ist. D.h., dass grundsätzlich jedem Begriff nur eine einzige Benennung zugeordnet ist, und umgekehrt. (...) Es sollte also weder mehrdeutige Benennungen (Homonyme und Polyseme), noch Mehrfachbenennungen für einen Begriff (Synonyme) geben.

(Wüster [1979] 1985, 79)²¹

In Wüster’s terminology (*Terminologielehre*), the starting point is the concept, not the denomination (Wüster [1979] 1985, 1). Wüster’s *Terminologietheorie*, as Oeser and Budin (1999, 2174) explain, even asserts that the realm of concepts in Terminology must be viewed as independent from the realm of denominations. In other words, science is a *Begriffssystem*. According to Wüster, there should be neither homonymy nor synonymy. For him, synonyms not only burden the memory, but are also confusing. This is especially true for the beginner in a subject field or in a language, who might come to believe that different denominations correspond to different concepts (Wüster 1970[1931], 97; Morgenroth 1990, note 139). Real synonyms should thus be eradicated in standardization (*Ausmerzung aller Synonyme*) (Wüster [1974] 1993, 348). Budin (1998, 178) sums up Wüster’s leitmotiv, which is “unambiguous communication in the professions and sciences”, and underlines Wüster’s main topics, namely, “terminology standardization and its basic principles”.

As previously mentioned, Wüster was actively engaged in official standardization work (ISA, ISO, ISO TC37):

20. the inadequate separation of the conceptual from the expressive side (our translation).

21. [One] must require terminology to be, to use an expression borrowed from mathematics, biunivocal, that is to say that, basically, each concept is assigned only one denomination, and vice versa. (...). There should therefore not be denominations with different meanings (homonyms and polysemes) or multiple denominations for one concept (synonyms) (our translation).

Die terminologische Gemeinschaftsarbeit darf sich nicht darauf beschränken, festzustellen, was ist. Sie muss danach trachten, den so widerspruchsvollen Sprachgebrauch zu vereinheitlichen und zu verbessern. (Wüster [1979] 1985, 89)²²

In order to achieve language standardization, unification and stabilization are necessary. Technical language standardization is about the standardization of technical products, as reflected in the work of ISO committees, ISO/TC37 and IEC, the International Electrotechnical Committee.

Already in 1931, Wüster wrote *Internationale Sprachnormung in der Technik* (...). This was the starting point for modern terminology work as well as for international standardization (ISA and ISO). He strove to harmonize international terminology while working together with other countries to establish international norms. He claimed that standardizing individual denominations needed unified and cross-linguistic guidelines from the *Allgemeine Terminologielehre* or terminology in general. Thanks to this type of action, many basic ISO standards in Terminology and Lexicography were developed in the space of twenty years by 43 countries and 35 organizations, including the UNESCO.

There are some other important elements concerning language development, related to standards (Wüster [1974] 1993, 339). In terminology work for national and international standards, concepts as well as denominations become internationally standardized. Unification of the *Sinnform* (*innere form*) is usually performed by unifying the basic meanings of combined elements and transferred denominations. Wüster was critical of the use of terms that spontaneously emerge from the chatter of the workshop, and which need to be regulated by standardization (Humbley 2018b, 161).

While many of Wüster's statements have become acknowledged principles, there is evidence of a less savory side since at least one of his statements is at best an ambiguous lapse in taste, and at worst, a reprehensible political statement. For example, on receiving an award from the Austrian Standards Institute, Wüster gave his acknowledgments (Typescript, 26/06 and 27/08/1970) the following title: "Normung macht frei"²³ which is reminiscent of the infamous inscription on the entrance gate of Dachau, Auschwitz and other Nazi concentration camps. However, quite surprisingly, the rigorous terminologist and standardizer, Wüster, showed a different attitude toward language variation.

22. Collaborative terminology work must not be limited to ascertaining what is. It must strive to standardize and correct the contradictory usage of the language (our translation).

23. Standardization sets you free (our translation).

5. Variation in Terminology

Wüster mentions certain situations where terminological stability might have to be relaxed:

Auch in der Terminologie muss das Verlangen nach vollständiger Eineindeutigkeit ein frommer Wunsch bleiben. (Wüster [1979] 1985, 79)²⁴

This is at variance with the principles advocated up to that point. Synonymy, distinctions between general language and terminology, diachronic or cross-domain usages, and evolving and unstable concepts are underlined or analyzed by Wüster. They even figure in his own terminology (see Candel 2004, 2007, 2017).

There is some variation in translating ‘Terminology’ (meant as a discipline: *Terminologielehre*). Oeser and Budin (1999, 2178) comment on Wüster’s *allgemeine Terminologielehre*:

‘Allgemeine Terminologielehre’ wie sie von ihrem Begründer Eugen Wüster genannt wurde (im Englischen übrigens – etwas aussagekräftiger – ‘General Theory of Terminology, – ein direkter (expliziter) oder indirekter (impliziter) Ausgang- oder zumindest Bezugspunkt für fast alle bekannten Theorien zu terminologischen Themen’.²⁵

The term *General Theory of Terminology*, or *Allgemeine Terminologielehre*, represents Wüster’s General Terminology. Even though his countless research works resulted in a theory, one should remember that for him it was not meant to be ‘A General Theory’ or ‘The General Theory’. Indeed, in an addendum to Wüster (1971a), Felber (1981b, 274), uses the term *Theorie der Terminologie* (theory of Terminology) when mentioning Wüster’s *Allgemeine Terminologielehre*. This may have helped to transform the ‘Science of Terminology’ into ‘The Theory of Terminology’ or ‘The General Theory of Terminology (GTT)’.

Variation may occur because of synonymy or homonymy. Synonymy as well as homonymy may be acceptable for Wüster, though in a measured way (see also Kocourek 1991, 187). One example is the case of *Kochsalz* (table salt) and *Natriumchlorid* (sodium chloride), which designate the same substance while referring to different *Sinnformen* (conceptual forms) (Wüster [1979] 1985, 84). Wüster even

24. Even in Terminology, the need for perfect biunivocity must remain wishful thinking (our translation).

25. *Allgemeine Terminologielehre* as it was called by its founder, Eugen Wüster (in English, the ‘General Theory of Terminology’, which is a bit more meaningful), is a direct (explicit) or indirect (implicit) starting point or at least a reference point for almost all known theories on terminology topics (our translation).

mentions synonyms in his own terminology: "Statt 'Logische Beziehungen' sagt man auch 'Abstraktionsbeziehungen'" ([1979] 1985, 9).²⁶ As previously mentioned in Section 4, "instead of 'general terminology' one can also say 'basic terminology':"

Selbst in der Terminologie können Synonyme mit Sinnformunterschieden nicht ganz unterdrückt werden. ([1979] 1985, 9)²⁷

Trotzdem muss man z. B. mit Synonymen unterschiedlicher Stilhöhe rechnen, wie sie die technische Hochsprache (das ist die Sprache der Normen) von der Werkstattssprache unterscheidet. (Wüster [1979] 1985, 84)²⁸

Wüster ([1959/1960] 2001 and [1959] 1963) wonders about the limits of standardization and biunivocity and discusses polysemy and synonymy. He is concerned about pairs of terms in his own metalanguage such as *Benennung* and *Terminus*, *Eigenmerkmale* and *inherente Merkmale*, *individueller Gegenstand* and *Individuum*, *Disjunktion* and *Aufpaarung*, *Integration* and *Anpaarung*, *Konjunktion* and *Abpaarung*.²⁹ Some terms are even reformulated twice, such as *quasisynonym*: *Ungefährsynonym*/ *Fastsynonym*/*Quasisynonym*). Wüster is not seeking to eliminate synonyms but only to minimize their use (Wüster [1979] 1985, 84). And interestingly he likes to use synonymy to facilitate defining (Candel 2007).

Variation is also present in specialized dictionaries. For example, Wüster's dictionary (1968), which improves the principles of Terminology, contains many synonyms. Standardization cannot be perfectly purist either. The development of international standards and the creation of new terms cannot always be a puristic endeavor since the same concept may be expressed by a German word as well as by a more international one (Wüster [1974] 1993, 348).

Even in standardization, when working in the ISA Committee TC37, Wüster explained that a synonym proposed by one member was taken into account by the whole committee (1951, in Bühler 1982a, III, 82). He used various ways of rewording, as mentioned in Candel (2007). For example, he resorts to implicit and explicit ways, objective and subjective ones, and sometimes to reformulations. This

26. Instead of "logical relations", one may also say 'abstraction relation' (our translation).

27. Even in terminology one cannot completely do away with synonyms and with formal semantic differences (our translation).

28. Nevertheless, one must for example take into account synonyms belonging to different stylistic levels, as the technical 'Hochsprache' language (formal technical language) distinguishes them from the language of the workshop (our translation).

29. Denomination and term, intrinsic features and inherent features, individual object and individual, disjunction and separation, integration and coupling, conjunction and pairing (our translation).

is accomplished either by simplifying the terms or specializing them, creating new terms or, eventually, trying to germanize them. Based on these considerations, it appears that Wüster's innovative work lies in making Terminology a new discipline though with a certain number of contradictions.

6. Wüster's position

Wüster often proposes relatively moderate options in terminology, not all of which belong to the same category. Although this might seem confusing, it shows how Wüster was always questioning himself. Wüster's shifting attitude when constructing his theories is reflected in his apparently contradictory positions regarding biunivocity:

Die Grundforderung für die Zuordnung zwischen Lautform und Bedeutung ist die der Eineindeutigkeit: Möglichst wenig Ausdrücke sollen mehrdeutig (Homonyme) oder mit anderen gleichbedeutend (Synonyme) sein. (Wüster [1931] 1970, 94)³⁰

Im allgemeinen ist absolute Eindeutigkeit (...) gar nicht erforderlich. Es genügt, wenn die Bezeichnungen im Zusammenhang eindeutig sind: relative Eindeutigkeit. (Wüster [1931] 1970, 95)³¹

Something similar occurs when he explains that *quality* is related to the *Ist-Norm* as well as to the *Soll-Norm* (Wüster 1971a in Bühler 1982a, II,131):

Ist-Richtigkeit: Übereinstimmung mit der gewählten sprachlichen Ist-Norm (...).³²

Soll-Richtigkeit: Übereinstimmung mit einer sprachlichen Soll-Norm (...). Diese ist entweder autoritär (...) oder vereinbart (...).³³

Even though Wüster was used to writing and rewriting his analyses and definitions (Bühler 1982a), he admitted that there is no easy decision when choosing a word or term, or when defining a discipline, and he clearly exemplified his hesitations.

30. The basic requirement for the assignment between phonetic form and meaning is biunivocity: As few expressions as possible should be ambiguous (homonyms) or have similar meanings to others (synonyms) (our translation).

31. In general, absolute biunivocity is not required at all. It is sufficient if the terms are unambiguous in context: relative univocity (our translation).

32. [Ist-] Correctness: Accordance with the selected linguistic Ist-Norm (...) (our translation).

33. [Soll-] Accuracy: Accordance with a linguistic Soll-norm (...) which is either authoritative (...) or agreed upon (...) (our translation).

In a letter (to Franz de Clavé, 8/01/1972, Typescript) concerning the names of the AILA sub-committees for Terminology and Lexicography, Wüster gives an interesting example of his way of analyzing words or terms. He even makes an initial proposal that does not seem quite logical. It is difficult to understand – much less agree with – what Wüster means by using the word *Terminologielehre* at the end of the following excerpt:

Das Wort Terminologie hat zwei abstrakte Bedeutungen:

(1) (...) “fachlicher Wortschatz”, d.h. fachliches Begriffs- und Benennungssystem.

Damit können entweder die Benennungen (= Termini) eines einzigen kleinen oder großen Fachgebietes gemeint sein oder die Gesamtheit der Benennungen aller Fachgebiete (ein- oder mehrsprachig).

(2) (...) “terminologische Theorie”, “Wissenschaft von den terminologischen Grundsätzen”, “Terminologielehre”. Die Terminologielehre ist ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften.

Die Terminologielehre ist überfachlich, d.h. nicht auf einen einzigen Sachbereich beschränkt.

Anmerkung zu (2): In dem deutschen Ausdruck “Terminologielehre” bedeutet “Lehre” nicht “Unterricht” (F enseignement), sondern “Theorie”. Das deutsche Wort “Lehre” ist bekanntlich zweideutig.³⁴

One may disagree with Wüster on point (2) since *lehre* means *science* and not *theory*. However, in a letter written four days later, Wüster corrects himself with a noticeable shift in his interpretation of *Terminologielehre*:

Ich möchte meinen Brief vom 8. Januar ergänzen.

Zu 2) Die Bedeutungen des Wortes “Terminologie” (...)

(2.1): Spezielle Terminologielehre, spezielle terminologische Lexikologie (...);

z. B. die deutsche medezininische Terminologielehre.

Diese Bedeutung habe ich in meinem Brief vom 8. Januar nicht erwähnt. (...)

(2.2): Allgemeine Terminologielehre, d.h. “Terminologie” im Sinne (2) meines Briefes vom 8. Januar, allgemeine terminologische Lexikologie.

Die allgemeine Terminologielehre ist nicht etwa die Summe der speziellen Terminologielehren (...), sondern sie ist durch Abstraktion [manually underlined] aus

34. The word *terminology* has two abstract meanings:

(1) (...) “technical vocabulary”, i.e. technical conceptual and naming system.

This can mean either the denominations (= terms) of a single small or large subject area or the entirety of the terms of all subject areas (monolingual or multilingual).

The Terminologielehre is interdisciplinary, i.e. not limited to a single subject area. Comment on (2): In the German expression, “Terminologielehre” means “Lehre”, not “teaching”, but “theory”. The German word *Lehre* is known to be ambiguous (our translation).

den speziellen Terminologien (...) gewonnen. Sie ist überfachlich (überdisziplinär) und übersprachlich. Sie steht auf einer höheren Abstraktionsstufe.

Die Begriffe (2.1) und (2.2) zusammen bilden die Terminologielehre oder terminologische Lexikologie.³⁵

The word *theory* is absent and Wüster now explains that his *Allgemeine Terminologielehre* (or *General Theory of Terminology*) is part of Terminology or of Terminological Lexicology.

An example of how theories emerge pertains to the title of Wüster's lectures and research. When Wüster was working with Manfred Dressler on the framework of his lectures at the University of Vienna (1972–1974), dealing with “Lexikologie, Lexikographie, besonders Terminologielehre und Sprachnormung”,³⁶ he proposed *Terminologielehre und Lexikographie* rather than *Sprachnormung, Terminologielehre* (Standardization, Terminology) (being mostly *fachliche Lexikologie* (Specialised Lexicology)) (1971a). He remarked that *Language Sciences* have much in common with *Terminologielehre* and then considered *Einführung in die Grundsätze der Terminologie und der terminologischen Lexikographie*,³⁷ and expressed his preference for *Lexikologie und Lexikographie*.

Over the years, Wüster considered a number of titles such as *Allgemeine Terminologielehre*, *AT*, *ATL*, *Allgemeine Terminologische Lexikologie*, *einfache Terminologiekunde*, *einzelfachliche Terminologielehre*, *Terminologielehre*, *Terminologische Grundsatzlehre*, *Theorie der Terminologie*, *Terminologielehre*, and *General Theory of Terminology*. In a letter to Wolfgang Dressler (1973), he explained that he wished to name these contributions *Grundbegriffe der Sprachgestaltung besonders der Allgemeinen Terminologielehre und der terminologischen Lexikographie*,³⁸ which he

35. I would like to complete my letter of January 8th. To 2) The meanings of the word “terminology” (...)

(2.1): Special terminology, special terminological lexicology (...); for instance German medical terminology.

I did not mention this meaning in my letter of January 8th. (...)

(2.2): General terminology, i.e. “Terminology” in the sense (2) of my January 8th letter, general terminological lexicology.

The General Theory of Terminology is not the sum of the special terminology teachings (...), but is obtained by abstraction from the special terminologies. It is interdisciplinary and cross-linguistic. It is on a higher level of abstraction.

The concepts (2.1) and (2.2) together form the terminology or terminological lexicology (our translation).

36. Lexicology, Lexicography, especially Terminology and Standardization.

37. Introduction to the Principles of Terminology and Terminological Lexicography.

38. Basic Concepts of Language Design, especially General Terminology and Terminological Lexicography.

corrected to *Grundbegriffe der bewußten Sprachgestaltung (...)* [basic concepts of conscious language formation].

The *Allgemeine Terminologielehre*, a new discipline coined by Wüster, was first presented in 1972 in Copenhagen as *Die Allgemeine Terminologielehre – Ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften*.³⁹ Some of Wüster's experts explain how his statements are to be understood. Oeser and Budin (1999, 2172) write:

die Forschungen zu einer allgemeinen Terminologielehre von Eugen Wüster, die als eine "Theorie der Terminologie" oder als "terminologische Grundsatzlehre" zu verstehen sind.⁴⁰

In his writings, Wüster shows many facets of himself as an experienced researcher and terminologist. Various names have been proposed for his work by himself as well as by others. There is *Terminologielehre*, as well as its synonyms, quasi-synonyms, variants or equivalents in other languages (such as *Allgemeine Terminologielehre*, *allgemeine T.*, *General Theory of Terminology*, *GTT*, *Science of Terminology*, and *Theorie der Terminologie terminologische Grundsatzlehre*), which may have different meanings and interpretations. According to Oeser and Budin (1999) (see Section 6), the English version of *Allgemeine Terminologielehre* is *General Theory of Terminology*, which is "a little more meaningful". This English formulation is somewhat stronger than the German because *-lehre* is upgraded to *theory*.

Wüster tried to find a title for his dictionary of definitions. Various proposals appeared in his writings, which he mostly calls *Terminology*, *Dictionary*, *Guide*, *Concepts* or *Basic Concepts*. Despite his experience as a researcher and terminologist, he hesitated when it came to naming the subject that he had worked on for more than twenty years (Bühler 1982a II, 31 sq.):

- Terminologie der Terminologien (1955)⁴¹
- Wörterbuch des Benennungswesens (1955, 1956) des fachsprachlichen Benennungswesens (1955, 1956, 1963)⁴²

39. General terminology – a borderline area between Linguistics, Logic, Ontology, Computer Science and Object Sciences.

40. Research on a general theory of terminology by Eugen Wüster, which is to be understood as a 'theory of terminology' or as 'basic terminological principles' (...).

41. Terminology of Terminologies (1955)

42. Dictionary of the Naming System (1955, 1956) / of the Naming System in technical languages (1955, 1956, 1963)

- Leitfaden des Benennungswesens (1956) des fachlichen Benennungswesens (1963)⁴³
- Die Begriffe des Benennungswesens – Systematisches Definitionswörterbuch (1962)⁴⁴
- Grundbegriffe des fachsprachlichen Benennungswesens (Systematisches Wörterbuch) (1963) der (Allgemeinen) Terminologielehre (Systematisches Definitionswörterbuch (1963) der Allgemeinen Terminologielehre (1966) der Sprachgestaltung, besonders der Allgemeinen Terminologielehre und der terminologischen Lexikographie / der bewußten Sprachgestaltung,...) (1973) der Terminologielehre (1976)⁴⁵
- Systematisches Definitionswörterbuch der Grundbegriffe der Terminologielehre (1976)

Figure 1 shows that in the beginning, Wüster opted once for *Terminology* (1955), three times for *Dictionary* (1955–1963), and finally once for *Systematic Definition Dictionary* (1976). He called it *Concepts* once in between (1962), but his most quoted title, with five uses, is *Basic Concepts* (1963–1976). He also mentioned *Guidelines* twice (1956–1963).

These choices reveal the continuous hesitations of this terminologist, who otherwise knew how to be quite categorical. The domains that he illustrated are the following: *Terminologies* (once, in 1955), *Naming* (seven times, from 1955 to 1963), *Terminology* (as (*Allgemeine*) *Terminologielehre*, once, in 1963) and as *Allgemeine Terminologielehre* (once in 1966), as *Basic Concepts of Terminologielehre* (once, in 1976), finally as *Terminologielehre* (twice, in 1976).

Interestingly, Wüster's goal was to explain the basic concepts of *Terminologielehre* rather than the *allgemeine Terminologielehre*. Wüster dealt with 'terminology', and not necessarily with 'general terminology', nor with what would eventually become a *General Theory* or the General Theory of Terminology (GTT).

43. Guidelines of the Naming System (1956) / of the Naming System in technical languages (1963)

44. Concepts of the Naming System, Systematic Definition Dictionary (1962)

45. Basic Concepts of the Naming System in Technical Languages (Systematic Dictionary) (1963) / of the (General) Terminology (Systematic Definition Dictionary) (1963) / of the General Terminology (1966) / of Language Design, especially General Terminology and Terminological Lexicography / of conscious speech formation (1973) / of Terminology (1976) / Systematic Definition Dictionary of basic Concepts in Terminology (1976).

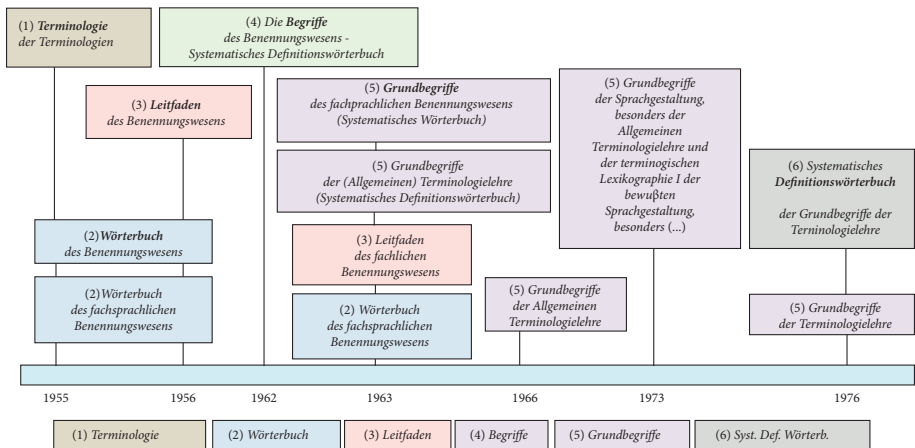


Figure 1. Timeline in Wüster's naming of his ongoing project

7. Final interpretation of Wüster's position

At this stage, it is necessary to revisit some interpretations of Wüster's work. Some authors may have misunderstood Wüster perhaps because they did not read his writings in German (see Humbley 2004; chapter in this volume). According to Romary *et al.* (2006), Wüster did not present monosemy as a dogma or deny variability. They underline the need for terminology standardization in addition to text-based terminology.

When L'Homme (2004, 24) mentions that the principles of the classical theory of terminology (or the General Theory of Terminology) are associated with Wüster, she indicates that what has been referred to as the General Theory of Terminology may simply be understood as the principles of the Classical Theory of Terminology.

Similarly, Cabré (2003, 166–168) recalls that if Wüster's followers went on to name his model the *General Theory of Terminology*, Wüster finally came to envisage his *Terminologielehre* as an autonomous interdisciplinary field of study. His *Terminologielehre* concedes a certain degree of synonymy (see also Candel 2004, 2007, 2017).

Rather than mentioning *THE General Theory*..., it would be suitable to refer to his work as Wüster's *Theory of Terminology*, or just Wüster's *Terminology* (which clearly contributed to positive developments, such as controlled synonymy or phraseology). Cabré (2003, 187) summarizes three perspectives on the specialized knowledge unit: (i) the cognitive one (the concept); (ii) the linguistic one (the term); (iii) and the communicative one (the situation). Antia (2002) considers Wüster to

be the “father of modern Terminology”, and describes him as a “pioneer of LSP who also tried to claim a place within Linguistics for LSP”, and who created the link between “terminology (standardization, synonymy, etc.) and the importance of diachrony” (see also Trojar 2017).

Despite the frontiers between different trends in Terminology, such as prescriptive or descriptive ones, these differences may eventually be evened out within terminology activities. Noting that rules and principles were already in use for emerging terms in special subject areas, Oeser and Budin (1999, 2171ss) observe that some general research on terminology was carried out at the beginning of the 20th century, in connection with standardization efforts. At that time, special terminological principles were being distinguished from general ones, the first of which focused on principles and methods for specific domains, and the second of which dealt with general principles and methods for all domains and for many languages:

Die normative Terminologearbeit setzt daher die deskriptive Terminologearbeit voraus. (2171)⁴⁶

Wüster ([1931] 1970) is more nuanced and subtle in his point of view than one may think. He points out the importance of both the *Bequemlichkeitspol* in everyday language – which Humbley (1989, 98) renders as *acceptabilité sociale des termes* [social acceptability of terms] – and the *Genauigkeitspol* [accuracy pole] of the scientific mode of expression. It is nevertheless quite logical that in some cases, these two aspects unite in the tendency or need for *Eineindeutigkeit* [biunivocity, unambiguity], thus linking the social acceptability of terms and biunivocity. Indeed, Wüster, as an expert in Terminology Theory, invites us to take linguistic variation into account. In this regard, the tendency toward standardization and excessive systematization is modulated by his awareness that a strong terminology is wishful thinking.

8. Conclusion

Throughout his life, Wüster was active in many areas, and engaged in numerous projects and international spheres. He worked tirelessly to express himself at various levels and left behind many documents, both published and unpublished. His contributions may not have initially attracted the international attention that they undoubtedly deserved, and sometimes he was criticized without having been read (see chapter by Humbley in this volume). However, he unquestionably contributed

46. Work in normative terminology therefore requires work in descriptive terminology.

to the identification of various facets of Terminology, which were of fundamental interest to terminologists, international standards experts, linguists, sociolinguists, and lexicographers. A crucial issue is that of prescriptivism or descriptivism. As reflected in his writings, his attitudes toward terminology shifted and oscillated, and this is probably what really constructed the field of Terminology.

He showed that in scientific or technical activities, one cannot do without prescription and biunivocity, and even more so at the international level. Clearly, when it comes to highly technical domains, it is necessary to get as close as possible to the concept in order to be internationally understandable or to at least be on an equal level with denominations in other languages. When referring to a technical object or concept, in a technical and international environment, strict terminological rules or ways of denominating may be required – and here the *Soll-Norm* should be imposed.

Considered to be the father of the General Theory of Terminology, Wüster effectively laid the groundwork of Terminology as a science. His *Grundbegriffe der Terminologielehre*, continuation of his 1931 *Internationale Sprachnormung...*, would have been the crowning glory of his endeavor (Letter to Kučera, 15/11/1976 (Bühler 1982a II, 41)). In this period a radical opposition grew up between languages for special purposes and general language, between concepts and denominations.

This chapter has underlined some rich though often contradictory positions in Wüster's work and statements. In his presentation of Terminology, he showed that one cannot do without prescriptive attitudes or descriptive ones either. Even though he shifted between radical positions and more moderate ones, he remains the founder of modern Terminology. His famous *Allgemeine Terminologielehre* is a Terminology – and in our opinion, it should not always be regarded as a theory, and may eventually be regarded as a discipline and as a science.

Acknowledgements

The author thanks the editors for their careful reading and helpful suggestions.

PART 2

Knowledge-based Terminology

Conceptual relations

From the General Theory of Terminology to knowledge bases

Anita Nuopponen

University of Vaasa

This chapter gives an overview of concept relations. They can be used to create terminological products, guarantee their quality and make their contents more accessible to the users. They facilitate the analysis, organization, and definition of concepts, and help to establish connections between concepts and terms, to form and evaluate terms, and to identify term equivalents. This chapter first relates conceptual relations to concept formation, and discusses relevant aspects of terminological resources. Then an extensive typology of conceptual relations is compared to typologies with multidisciplinary backgrounds and purposes. Terminology work and terminological resources often use only basic relation types. However, ontologies, automatic term extraction, and advanced terminological knowledge systems benefit from a larger set of relation types.

Keywords: Terminology, concept relation, relation, relation typology, semantic relation, concept, terminology work, terminology research

1. Introduction

In terminology work, concept relations are important to create terminological products and to make their contents accessible to the users. They are instrumental in the extraction of conceptual information, the analysis and organization of concepts, the definition of concepts, the specification of connections between concepts and terms, and term formation and evaluation. Furthermore, they are needed when presenting terminological information in vocabularies, databases, knowledge representation systems such as ontologies (see chapter by Montiel-Ponsoda in this volume), and other resources such as terminological knowledge bases (see chapter by Meyer in this volume). They guarantee the quality of terminological products and facilitate knowledge transmission and acquisition.

The identification of relations between concepts and their organization in concept systems was a central element of the General Theory of Terminology developed by Eugen Wüster (see chapters by Candel and Humbley in this volume). It was the basis for compiling vocabularies for special fields, and was introduced as principles for terminology standardization (see chapter by Wright in this volume). What makes the role of concept relations even more relevant today is, on the one hand, the need for terminology work in the digitalization and, on the other hand, the development of next-generation terminological products and conceptual knowledge systems as well as information extraction applications that are still on the horizon. In the 1990s, Meyer and her colleagues were looking forward to the next generation of terminological repositories that would contain “a richer and more systematically structured knowledge component than do conventional term banks and specialized dictionaries” (Meyer, Eck and Skuce 1997, 99; see also Skuce and Meyer 1990 and chapter by Meyer in this volume). However, this task has taken some time to develop and accomplish.

This chapter reviews sets of concept relation types that can be used to develop terminological resources and to define methods to compile them, e.g., domain and task-specific relation typologies. This chapter is organized as follows. Section 2 relates concept relations to the process of concept formation. Section 3 discusses certain aspects of terminology resources where concept relations are relevant. Section 4 analyzes relation types based on Nuopponen (1994a, revised in 2005) to build on terminological literature and incorporate ideas from related studies. Section 5 presents the conclusions that can be derived from this chapter.

2. Concepts, characteristics and basics about concept relations

In Terminology, concepts are often defined as abstract units of thought or units of knowledge, consisting of characteristics (Figure 1), which correspond to one or more objects. Characteristics are abstractions of properties that are found or assumed to exist in real or imagined phenomena or objects (of reference) or which appear from their relationship with other phenomena.

Sager (1990, 22) describes concept formation as a “process of variously grouping and ordering the material and immaterial objects which we sense, perceive or imagine into abstract categories”. Concepts thus become “constructs of human cognition processes which assist in the classification of objects by way of systematic or arbitrary abstraction” (*ibid*).

Figure 1 depicts the three levels, which are relevant to terminology work and research: (i) phenomena or objects (real world); (ii) ideas and concepts (cognition); and (iii) expression and communication (communication). The model is a simplification as many of the objects of reference exist only in people’s minds.

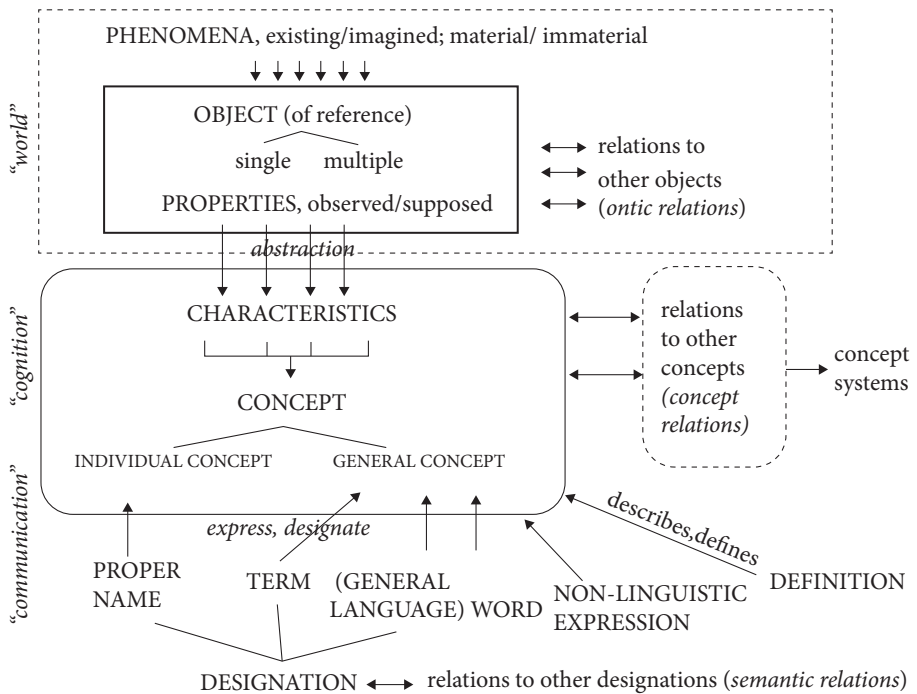


Figure 1. Concept formation

Abstraction leads to concept formation, which can be taken to an even higher abstraction level based on the shared characteristics of two or more concepts. This results in generic concept relations and systems (see Section 4.1). Moreover, in many cases, objects can be partitioned or segmented in phases or causes (Nuopponen 2007, 204). The concept relations stemming from the abstraction of these 'real life' connections can be classified as ontological concept relations (Wüster [1974] 1994, 95; Nuopponen 1994a, 84, 2018a, 457) (see Section 4).

Concept relations can be defined simply as relations between concepts (ISO 1087:2019, 4). They are different from relations between terms, i.e., linguistic designations for specialized concepts (see chapter by L'Homme in this volume) and from the relationships between terms and concepts. These distinctions are not always made, especially in multidisciplinary terminological studies. Often, the term *semantic relation* is used as a synonym for *concept relation*. This chapter focuses on relations between concepts, and thus the preferred term is *concept relation*. Relations are viewed here from the conceptual perspective regardless of which term is used in the sources.

3. Concept relations in terminological resources

As previously mentioned, concept relations are important elements in terminology work and compiling terminological resources. The development of new generations of terminological knowledge bases emphasizes the need to organize entries and knowledge, and to implement concept relations in term bases.

In systematic terminological vocabularies, entries are organized thematically or according to concept relations as in the vocabulary of terminology work ISO 1087:2019. In addition, the entries can be numbered in a way that make the hierarchical organization of concepts explicit. Concept relations can also be indicated in diagrams or tables depicting concept systems, information in definitions and explanations, references between entries, systematic lists in indexes, etc. (Nuopponen 2018a, 456).

Meyer et al. (1992), León-Araúz, Reimerink and García-Aragón (2013, 31) and Faber (chapter in this volume) emphasize that terminological knowledge bases should reflect the way that concepts are related to each other in the human mind. According to Meyer, Eck and Skuce (1997, 105), it is “very unfortunate that representations of concept systems are still quite rare in specialized dictionaries and terminological databases”, something that is true even today. This is the kind of laboriously acquired subject-field knowledge, which Meyer, Eck and Skuce (1997, 99) see as valuable for users, but which generally “stays where it was first stored, namely in the terminologist’s head” with only fragments included in definitions and examples. Marshman, Gariépy and Harms (2012, 47) agree that “the benefits of including terminological relations in many cases will outweigh the modestly increased workload, and that (as is the case with translation memories) the gradual accumulation of information will ultimately form a useful resource.” What is needed is a termbase tool with a “structure that is adequate for storing the information and providing quick and multifaceted access” (ibid, 48).

One of the major benefits of making concept relations visible in terminological products is to enable knowledge transfer. In accordance with Meyer, Eck and Skuce (1997), Faber (2011, 10) states that “knowledge of conceptualization processes as well as the organization of semantic information in the brain should underlie any theoretical assumptions concerning the access, retrieval, and acquisition of specialized knowledge as well as the design of specialized knowledge resources.” According to León-Araúz and Faber (2010, 17) the system thus becomes “a consistent resource in its different representational levels” because contextual constraints “structure knowledge in a similar way to how things relate in the real world, as well as in the human conceptual system”.

Concept relation information in terminological resources also has a didactic value (Picht and Draskau 1985, 92) since it helps the user to better understand terms, concepts, and their definitions (Nistrup Madsen, Sandford Pedersen and Erdman Thomsen 2001, 7). Understanding concepts in a specialized field helps language professionals to become familiar with a new domain and its language (Marshman, Gariépy and Harms 2012). Concept relation information and the visualization of concept systems also enable browsing without specific knowledge of a term or its exact orthographic form. This is evidently not the case of alphabetically organized glossary entries or a termbase interface with a simple search window.

4. Types of concept relations

Concepts can be related to each other in many ways, which is accurately explained by Sager (1990, 29):

Inside subject fields concepts are also related either by their nature or by the real-life connections of the objects they represent. As in real life between objects, the kinds of relationships which exist between concepts are numerous and varied.

There are not only numerous types of concept relation, but also various ways to define and classify them. How concept relation types are distinguished from each other depends on the theoretical background or discipline and purpose or context of use or need (Nuopponen 2014, 2). This section shows how typologies can differ from one project to another. Firstly, Figures 2–4 present a concise overview of the proposals in Nuopponen (1994a, revised in 2005).¹

The relations in our typology are defined and complemented with relation types from three multidisciplinary terminology projects (Table 1). These projects base their relation inventory on the empirical analyses of relations in specific domains. The typology in Figures 2–4, in contrast, seeks to create an overall classification

1. The figures include corresponding or similar relation types from the other typologies. The main distinction is made between generic (logical) and ontological relations, which cover all the other relation types. In the figures, the node *ontological relations*, which would cover all except generic relations, has been left out. The typology also integrates and builds on relation types from less recent terminological literature, e.g., Wüster (1971b, 1974 [1993]), Dahlberg (1978, 1981), DIN 2330-1979; DIN 2331-1980; Arnzt and Picht (1989a) as well work in Philosophy, Linguistics, and Semantics. The typology functions as the core of Systematic Concept Analysis (see Nuopponen 2011), a method that includes also a satellite model in the form of a mind map as a tool for compiling and organizing concepts with the help of various types of relation (see Nuopponen 2016).

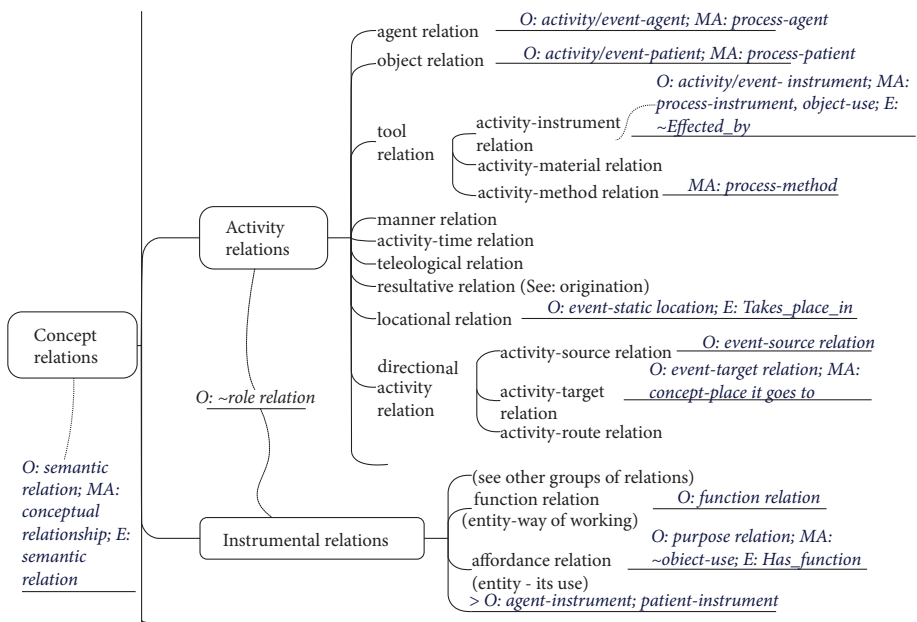


Figure 2. Concept relation typology: Generic and contiguity relations (Nuopponen 1994a, 2005)²

of concept relations that would be applicable to all domains and languages. We are aware that some of these relations types overlap with the ones presented in Figures 2–4, but our descriptions of specific relations (Sections 4.1 to 4.7) also refer to these typologies.

Table 1 lists the relation types of the other typologies and the different labels used in them. Table 1 first shows the 2004 version of the OntoQuery typology.³ The OntoQuery (ontology-based querying) project described by Nistrup Madsen et al. (2001) integrates terminological methods in ontology work, why their relation typology combines relations used previously in both terminology (e.g., Nuopponen 1994a) and ontology work. Consequently, they use the terminology from both fields, while the term *semantic relations* for concept relations comes from the latter one.

2. Markings in the figures: O = OntoQuery, MA = Maroto and Alcina 2009, E = EcoLexicon; ~ = has similarities, > = to be included.

3. Other relations are discussed in other articles related to this project.

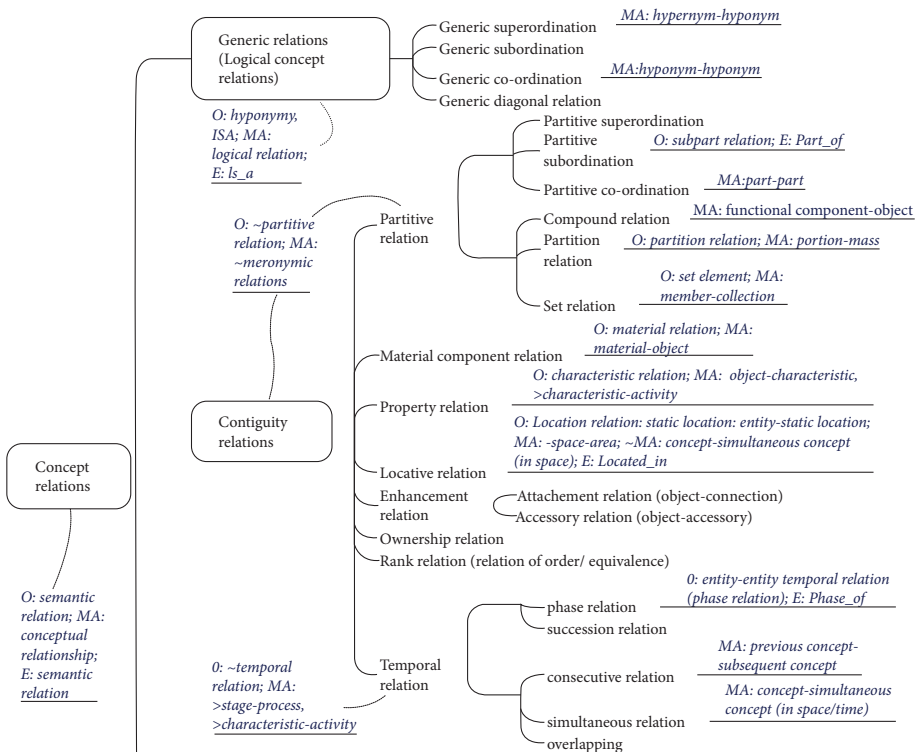


Figure 3. Concept relation typology: Activity and instrumental relations (Nuopponen 1994a, 2005)

Maroto and Alcina (2009) present a typology which is based on Sager (1990) and enhanced with the meronymic relationships in Winston, Chaffin and Herrman (1987). The authors are interested in how concept relations should be stored in a conceptual database so that this information can be used in knowledge retrieval. The third typology is the inventory of relation types from EcoLexicon-related research (e.g., León-Araúz and Faber 2010). Their typology is based on corpus studies where relations are often named with the most frequent or typical lexical pattern for each relation type, e.g., *is_a*, *part_of* (see chapter by Faber in this volume).

In Nuopponen (2005, 128–129, 2014), the upper-level hierarchy consists of logical (generic) and ontological relations as reflected in Wüster (e.g., [1974] 1993, 360–361, 1985, 9, 12). Terminological literature generally distinguishes between hierarchical and non-hierarchical relations. Generic and partitive concept relations are regarded as hierarchical and all the other types of relations as non-hierarchical (e.g., ISO 704:2009; ISO 1087:2019). Non-hierarchical relations are also referred

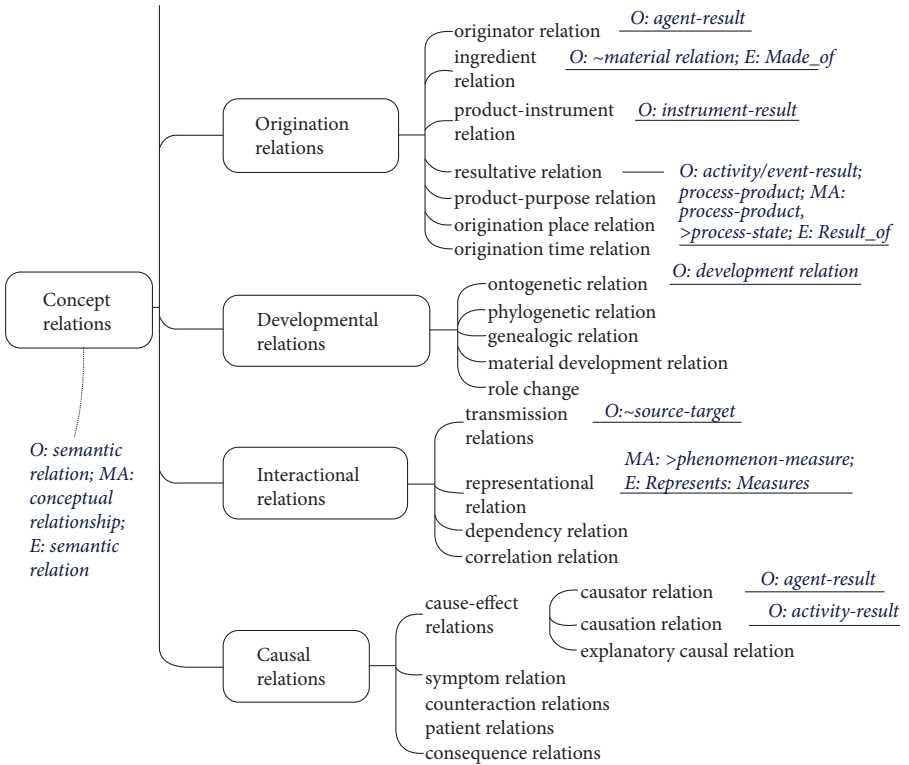


Figure 4. Concept relation typology: Origination, developmental, interactional and causal relations (Nuopponen 1994a, 2005)

to as *associative*, which is sometimes a synonym for *ontological relations*. However, they are not identical since associative relations include all ontological relations except for partitive ones (see Nuopponen 2014, 5–8). On the other hand, associative relations are sometimes referred to as *undefined relations*, when there is no matching relation at hand (Nistrup Madsen et al. 2001, 18; Maroto and Alcina 2009; Table 1). In the following, the relations are organized as depicted in Figures 2–4, where the upper-level node *ontological relation* has been omitted.

Table 1. Concept relation typologies

OntoQuery:	– Hyponymy (ISA)	relation; material relation;
Andreasen et al. (2004, 205):	– Location relation: dynamic location (source SRC: source-target /event-source relation; target relation: source-target /event-target relation); static location (event-static loc./entity-static loc.)	set-element
Semantic relations	– Causal relation	– Role relation: agent relation (event-agent/ agent-patient/ agent-theme/ agent-result relation/ etc.); patient relation; theme relation; instrument relation; result relation (All the roles are combined with each other)
	– Purpose relation	– Measurement relation
	– Event relation: event-source / event-target relation/ event-static loc./entity-static loc./ event-agent / event-patient/ event-theme/ event-instr. /event-result relation	– Characteristic relation
	– Function relation (entity-way of working)	– Temporal relation: time-entity; entity-entity temporal relation (phase relation; development relation)
	– Partitive relation: subpart relation; partition	
Maroto and Alcina (2009): Conceptual relationships	– Logical relations: hypernym-hyponym; hyponym-hyponym	concept-simultaneous concept (in time); previous concept-subsequent concept
	– Meronymic relationships: functional component-object; member-collection; portion-mass; material-object; stage-process; characteristic-activity; space-area; part-part	– Argumental and circumstantial relationships: process-agent; process-patient; process-product; cause-effect; process-instrument; process-method; object-use
	– Sequential relationships: concept-simultaneous concept (in space); concept-place it goes to;	– Other relationships: phenomenon-measure; object-characteristic; associative relationship
EcoLexicon:	Is_a; Part_of; Phase_of; Made_of;	Affects; Has_function; Effected
Semantic relations	Delimited_by; Located_in; Takes_place_in; Result_of; Causes;	by; Represents; Measures

4.1 Generic relations

In terminology literature, the most frequently mentioned relations are generic relations, also known as *logical relations*. According to ISO 1087:2019, generic relations are relations between a generic concept and a specific concept (e.g., ‘software’ – ‘application software’), where the intension of the latter includes the intension of the former and at least one additional delimiting characteristic. The generic concept

in the generic super-/subordination is called a *superordinate concept* whereas the more specific concept is the *subordinate concept*. The set of subordinate concepts on the same level of abstraction are known as *coordinate concepts* (e.g., ‘system software’ – ‘application software’ in Figure 5). (see ISO 1087:2019; ISO 704:2009; Nuopponen 2018b, 458–459, Table 2).⁴

Table 2. Generic concept relations

Generic (logical) concept relations	Relation participants	Examples
Generic super/subordination	superordinate (higher level of abstraction), subordinate concept (lower level of abstraction)	
a. direct generic super/subordination	super- and subordinate concepts on subsequent abstraction levels	‘software’ – ‘application software’
b. indirect generic super/subordination	super- and subordinate concepts on a different abstraction level further away	‘software’ – ‘text processing software’
Generic co-ordination		
a. direct generic co-ordination	co-ordinate concepts on the same level and below same direct superordinate concept and same criteria of division	‘application software’ – ‘system software’
b. indirect generic co-ordination	co-ordinate concepts on the same level, but under different direct superordinate concepts or criteria of division	‘application software’ – ‘server software’
Generic diagonal relation	other pairs of concepts on different abstraction levels in the same concept system	‘system software’ – ‘text processing software’

Members of the EcoLexicon project, for instance, examine an alternative way of distinguishing between different types of generic relations, based on the type of characteristic. Since EcoLexicon generated different types of generic relations at the same level, it was necessary to reduce noise, information overload, and redundancy (Léon Araúz et al. 2016, 76.). Gil-Berrozpe, León-Araúz and Faber (2016) analyzed *rock* and found the following subdivisions for specific concepts: formation-based, composition-based, location-based, state-based, attribute-based, function-based, and shape-based hyponymy.

This kind of specification of generic relations can be domain-dependent, and for each field, classification must be done separately. Instead of distinguishing between different types of generic relations, this is usually done by finding types of characteristics that function as criteria for subdivision (Bowker 1997a, chapter in

4. The counterparts in semantic relations are hyponymy/hyperonymy, hyperonym, hyponym and co-hyponyms (see Murphy and Koskela 2010).

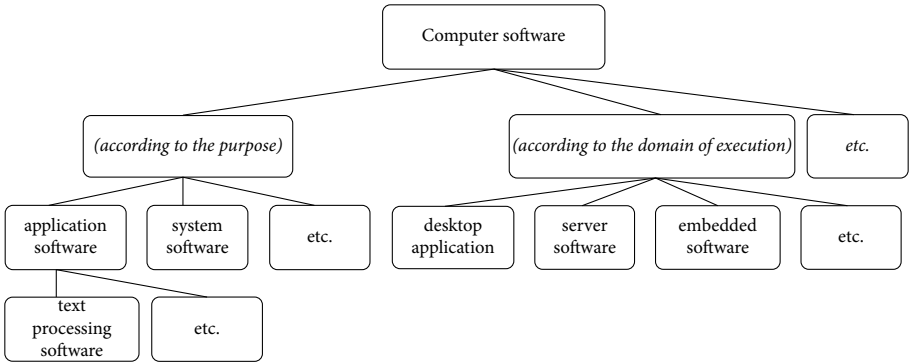


Figure 5. A fragment of a generic concept system for computer software

this volume; ISO 704:2009, 11–13; Nistrup Madsen and Erdman Thomsen 2008, 3). Subdivision is thus based on characteristics, (e.g., size, origin, purpose, formation, composition, location, state). For instance, computer software can be organized according to purpose (e.g., ‘application software’ and ‘system software’) or to the domain of execution (e.g., ‘desktop application’, ‘server software’, ‘embedded software’, etc.) (Figure 5).

4.2 Contiguity relations

Contiguity relations are dealt with here as a group even though they are different relation types. Although most of them are generally classified as partitive relations, here a more detailed classification is made (sse Table 3).

Table 3. Contiguity relations

Relations of contiguity	Based on relationships between	Example
partitive relations: a whole and its parts (whole-part, part-part)		
– compound relation	a whole and its functional/constituent part	‘house’ – ‘door’
– partition relation	a whole and the part “where the nature of the part is the same as that of the whole”	‘cake’ – ‘slice’
– set relation	a whole and a group of similar things forming part of it	‘forest’ – ‘tree’
material component relation	a whole and components which can be extracted from it	‘wine’ – ‘alcohol’; ‘olive’ – ‘olive oil’

(continued)

Table 3. (*continued*)

Relations of contiguity	Based on relationships between	Example
property relation	an object and its property	'wool' – 'thermal resistance'
locative relation	an object and the place where it appears, is to be found in, is kept in (container – contained)	'furniture' – 'room'; 'auditory information' – 'echoic memory'
enhancement relation	an object and another object that can be attached to it, not an essential part (object – accessory / attachment)	'e-mail' – 'attachment'; 'car' – 'spare tyre'; 'camera' – 'USB cable'
ownership relation	an owner and something owned	'car owner' – 'car'
rank relation	objects that are evaluated and ordered according a certain type of property (order or equivalence)	'president' – 'vice president'; 'gold medal' – 'silver medal'
temporal relations:	actions or objects with contact in time (consecutive, overlapping or simultaneous)	
– phase relation	– phases/stages in a process/event/action	'drafting' – 'rewriting' – 'editing' etc.
– succession relation	– a series of persons or things that follow one after another	'starters' – 'main course'

4.2.1 Partitive relation

Partitive (concept) relations are based on dividing a whole into parts (e.g., 'car' – 'steering system'), which explains why they are sometimes called *part-whole relations*. In addition to relations between whole and part concepts, partitive relations include *part-part* relations that are based on parts of the same whole (e.g., 'steering wheel' – 'steering column'). Since they thus create hierarchies in a way that is reminiscent of generic relations, they are also classified as hierarchical (e.g., ISO 1087:2019; ISO 704:2009). Analogically, the concepts are often called *super-, sub- and co-ordinate concepts* in both of them even though the nature of the relations is totally different (ISO 704:2009, 8). The superordinate concept in partitive relations is also called *comprehensive concept* as opposed to *partitive concept* (ibid. 13).⁵

Based on Cruse's (1986, 162) division of canonical and facultative meronyms, Nuopponen (1994a, 161; 2005, 132) also distinguishes between canonical or facultative partitive super- and subordination in partitive concept systems. This means

5. The counterpart in semantic relations is called *meronymy* and the term is sometimes used to refer to partitive concept relations, too (Maroto and Alcina 2009, 237). This type of semantic relation has been dealt with quite extensively in Cruse (1986) and Murphy (2003).

that an entity needs the part (canonical superordination; e.g., ‘car’ – ‘tyre’) or can exist without the part (facultative superordination; e.g., ‘hotel’ – ‘restaurant’); the part cannot exist without the whole (canonical subordination; e.g., ‘finger’ – ‘hand’), or does not necessarily need a whole (facultative subordination; e.g., ‘tree’ – ‘forest’).” (Nuopponen 1994a, 161–164, 2005, 132; see also Winston et al. 1987).

Partitive relations can be divided into the compound relation (*sub-part relation*) (‘bicycle’ – ‘wheel’), partition relation (‘bread’ – ‘slice’), and set-element relation (‘firm’ – ‘employer’) (Nistrup Madsen et al. 2001, 7). Maroto and Alcina (2009, 236) define the partition relation (*portion-mass*) as a relationship “between a part and the whole where the nature of the part is the same as that of the whole”, (e.g., ‘specimen’ – ‘material’). The set-element relation (*member-collection*) is described as a relationship that is “established between parts that have space proximity or social connection with respect to the whole, no matter the function they fulfil or the fact that they are situated structurally in a specific way”, (e.g., ‘mosaic’ – ‘tesserae’ [piece of mosaic]) (ibid).

4.2.2 *Material-component relation*

Another type of contiguity relation is the relation based on an entity and a material that forms a part of it or can be extracted from the object, e.g., ‘wine’ – ‘alcohol’; ‘olive’ – ‘olive oil’; ‘egg’ – ‘cholesterol’. This relation is different from partitive relations and treated as a separate type of contiguity relations (Nuopponen 1994a, 91, 108–110; 2005) (cf. Nistrup Madsen et al. 2001, 7; Maroto and Alcina 2009, 236; see also the ingredient relation in Section 4.4). Nevertheless, this distinction is not relevant when analysing concepts that refer to abstract phenomena (e.g., ‘friendship’ – ‘trust’).

4.2.3 *Property relation*

The property relation is based on the relationship between an object and its properties, e.g., ‘wool’ – ‘thermal resistance’; ‘virus’ – ‘contagious’. The object can be of any type (e.g., material, activity, entity). The properties could also vary considerably (e.g., form, performance, purpose, color, ability etc.). Maroto and Alcina (2009, 236) exemplify the *object-characteristic* relationship with ‘ceramic floor tile’ – ‘water absorption’. In some domains, such as Chemistry and Law, property concepts are important. However, in other fields and contexts, properties are not always defined as concepts or given a designation, but rather specified with longer descriptions.

4.2.4 *Locative relation*

Locative relations are based on the contact between an object and its location, site, habitat, environment or container (e.g., ‘fish’ – ‘lake’/‘water’; ‘book’ – ‘bookshelf’; ‘tea’ – ‘teapot’; ‘visual information’ – ‘iconic memory’). León-Araúz and Faber (2010,

13) consider this relation type “relevant when the location of a physical object is an essential characteristic for its description”. They distinguish it from the partitive relation and exemplify the latter with ‘river bed’. Since a river cannot exist without a bed, there is thus a partitive relation. If these two relations converge, the *part_of* relation overrides the *located_in* relation (ibid). Another example is ‘computer case’ since the central processing unit (CPU) is located in the case. The CPU is thus not part of the case but both are parts of the computer configuration.

Maroto and Alcina (2009, 236) define *space-area* as a relation that is “established between a spatial area and the specific localizations inside that area”, e.g., ‘visible area’ – ‘radiation field’. Moreover, other relation type classes include location, linked to the relations of activity (Section 4.3), origination (Section 4.4), and transmission (Section 4.6). However, they involve a more dynamic location, a location where something happens.

4.2.5 *Enhancement relation*

The enhancement relation connects two objects, one of which can be attached to the other without being its actual part (Nuopponen 2005). It is not always distinguishable from (facultative) partitive relations. The question here is the difference between *part-of* (e.g., ‘mobile phone’ – ‘camera’) and *accessory-of* (e.g., ‘mobile phone’ – ‘protective case’). Dictionaries define ‘accessory’ as “an object or device that is not essential in itself but adds to the beauty, convenience, or effectiveness of something else” (Merriam-Webster 2020). A distinction could be made between relations referring to (a) an object and its accessories (accessory relation); and (b) an object and a connected object (attachment relation).

4.2.6 *Ownership relation*

The ownership relation is sometimes impossible to distinguish from the partitive relation, but it may become relevant in certain cases, such as in ‘car owner’ – ‘car’ and in other cases where ownership is an essential characteristic (e.g., ‘copyright owner’ – ‘copyright’; ‘building cooperative’ – ‘real estate’). The relation often involves concepts referring to persons, groups of people or legal entities, and an object belonging to them.

4.2.7 *Rank relation*

The rank relation is based on comparison, rating, or ranking phenomena according to a property, such as importance, efficiency, value, or size. A classic example is the different levels in the military, e.g., ‘sergeant’ – ‘corporal’ – ‘private’. Furthermore, concepts in a taxonomy may be related to each other with rank relations, e.g., ‘species’ – ‘genus’ – ‘family’ – ‘class’. Examples from academic education include

‘Bachelor of Arts’ – ‘Master of Arts’ – ‘Doctor of Philosophy’. The concepts on the same level are ‘ranked’ equally, e.g., ‘Master of Arts’ – ‘Master of Science’. The rank relation is the original hierarchical relation). Various fields have their own hierarchies and principles upon which hierarchies are built. They also have their own terminology as well as meta-concepts, which terminological methods could apply to further specify this relation type (Nuopponen 1994a, 94–95, 186–187, 2011, 13).

4.2.8 Temporal relation

Temporal relations are based on contact in time and should be distinguished from contact in space. Depending on the concepts connected, they could be subdivided into phase relations or successor relations. In the first case, concepts represent a sequence of phases of events, actions, or processes (e.g., ‘drafting’ – ‘rewriting’ – ‘editing’ etc.). In successor relations, the concepts refer to persons, things, or time periods that follow one another in succession, (e.g., ‘former president’ – ‘president’; ‘starters’ – ‘main course’; ‘Monday’ – ‘Tuesday’ – etc.). In both cases, these relations could be subdivided into consecutive, overlapping (intertwining), and simultaneous (Nuopponen 1994a, 176–182; 2007, 204). Therefore, they could also be classified as sequential as in ISO 1087:2019. In addition to the relations between each of the phases, and relations “between two activities or entities that concur”, Maroto and Alcina (2009, 236) distinguish relations between an activity as a whole and its phases. This can be compared to the spatial whole-part relation.

Temporal relations are an essential element in concept systems with relations of activity (Section 4.3), origination (Section 4.4), and cause (Section 4.7). For instance, in concept systems based on activity relations, temporal relations combine different activities such as the phases of a process or a multiphase activity. Figure 6 shows a sequence of three phases that represents the analysis of concepts for different phases of an activity or a process, different stages in the production of a product, or different “links” in a causal chain.

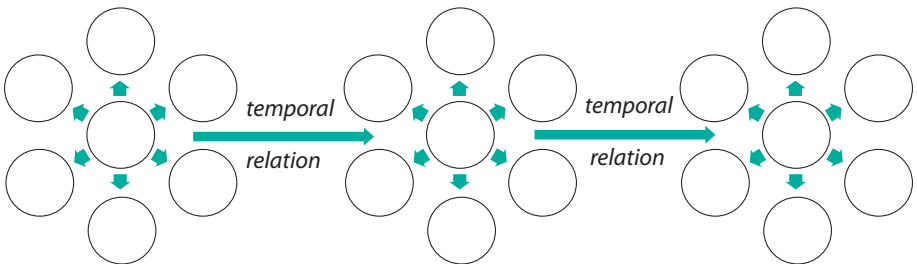


Figure 6. Temporal relations as connectors

4.3 Activity relations

Activity relations are a set of concept relations, where one of the related concepts represents an activity (e.g., 'fishing', 'gardening', 'elections', 'terminology work', 'meteorology'). The other related concept refers to the entity performing the activity (agent), the object of the activity, or the tools, materials or methods used (Table 4). Additionally, it may refer to the purpose, location, or direction of the activity. Each of them may further result in a broad generic concept system, such as tool concepts where a set of instruments, materials, methods etc. are listed and specified (Nuopponen 1994a, 105; 2006, 85–86; 2011, 11–12). Different sequential activities can be connected by temporal phase relations as shown in Figure 6.

The 'agent' concept refers to the entity that performs the activity. It can be "a person, a group, or an organization that intentionally carries out some activity or contributes to it, or takes part in it" (Nuopponen 2006, 85). In the EcoLexicon project, activity and agent also play a central role (e.g., Faber 2015, 23). Here a distinction is made between purposeful, goal-oriented action (agent relation) and a causal activity, where the 'causing agent' may often be a "natural agent" (see Section 5.7).

An 'object' in this context is something that is affected by the activity, or that undergoes an activity, such as 'fishing' – 'fish', 'interior design' – 'interior', 'nursing' – 'patient', or 'cognitive science' – 'cognition'. A 'tool' in this context is understood very generally as anything that is used to perform an activity or to carry out an action ('cleaning'), e.g., equipment (or instrument) ('mop'), material ('detergent'), and method ('green cleaning'). The manner of doing something can also be regarded as a tool or can be taken separately (Nuopponen 2006, 86–87, 2018b).

Instrumental relations could also be taken as a relation class of their own since apart from the activity and origination relations (Section 4.4), there are other relations involving instruments. Maroto and Alcina (2009, 236) include the *object-use* relation in their classification, where 'object' refers to the instrument and 'use' to the use or function "to which it is devoted". These relations could also encompass those referring to the possible uses of a tool (affordance). Faber, León-Araúz, and Prieto Velasco (2009, 11) include the *has-function* relation, which may link both artificially created objects or processes and natural entities "for human profit". Furthermore, they mention relations such as *functions*, *measures*, *studies*, *represents*, e.g., "a hydrograph represents rate of water flow" (ibid).

Another group of relations that has been taken as separate in the OntoQuery project are the role relations (Nistrup Madsen et al. 2001, 18–24; Andreassen et al. 2004, 205; see Table 1).⁶ In Table 4, the activity concept is the core concept and

6. The typology is partly based on the activity relations in Nuopponen (1994a, 105, 205; 2005), but includes some further relation types and takes a different angle to the concepts involved.

Table 4. Activity relations

Activity relations	Based on relationships between	Examples
agent relation	activity and its performer	‘research’ – ‘researcher’
object relation	activity and its object	‘research’ – ‘research question’
tool relation	activity and tools used for it: (instrument, material, method)	‘research’ – ‘research method’
– activity-instrument (equipment) relation	activity and instrument / equipment	‘research’ – ‘software’
– activity-material relation	activity and material utilized	‘research’ – ‘research material’
– activity-method relation	activity and method utilized	‘research’ – ‘qualitative method’
manner relation	activity and the manner/way to do it	‘research’ – ‘objectivity’
activity time relation	activity and time for it	‘research’ – ‘research leave’
teleological relation	activity and its purpose	‘research’ – ‘research result’
locational relation	activity and the place	‘research’ – ‘university’
directional activity relation	activity and directions for it:	
– activity-source relation	activity and its starting point	‘marathon’ – ‘starting line’
– activity-target/goal relation	activity and its end point	‘marathon’ – ‘finish line’
– activity-route relation	activity and its route	‘marathon’ – ‘marathon course’

the others are directly linked to it (see Nuopponen 2011, 11–12; 2006). However, from the perspective of role relations, all the concepts in a concept system of activity have connections between them (e.g., activity/event-agent, agent-patient, agent-instrument, and agent-result) (Andreasen et al. 2004, 205). This would apply more widely to all relation types since concepts can take certain roles (e.g., agent, patient), similar to semantic roles. The EcoLexicon knowledge base, for instance, applies roles, and relates each concept type to other concepts by a set of conceptual relations (Faber, León-Araúz and Prieto Velasco 2009, 8).

4.4 Origination relations

Concept relations of origination exist between concepts that refer to a concrete or abstract object and those that refer to its origin (Table 5). The main concept is the first one. Related concepts refer to the producer/maker/creator/etc., material or ingredients of the object, instrument, manufacturing method, process or activity, place of origin, and time of origin as well as purpose of the product (Nuopponen 1994a, 107–110, 205; 2011, 9).

A distinction is made between ingredient, material component and partitive relations. The ingredient relation is based on materials or ingredients that are used to make something, e.g., ‘bread’ – ‘flour’, ‘yeast’, ‘salt’ and ‘water’, whereas the material components (see Section 4.2.2) of bread would be those in its chemical or nutritional composition e.g., ‘bread’ – ‘gluten’, ‘acrylamide’, ‘carbohydrates’. The relation ‘bread’ – ‘crust’ is partitive (Section 4.2.1). Both artificial and natural objects can be analyzed as to the materials that they are made of (see Faber, León-Araúz and Prieto Velasco 2009, 10), which materials they can be reduced to, or which functional parts they can be broken into.

Table 5. Origination relations

Origination concept relations	Based on relationships between	Examples
originator relation	object and its originator, creator, producer	‘bread’ – ‘baker’
ingredient relation	object and ingredients it is made of	‘bread’ – ‘flour’, etc.
product-instrument relation	object and a tool or a method used to produce it	‘bread’ – ‘oven’
resultative relation	object and the activity of producing it	‘bread’ – ‘baking’
product-purpose relation	object and purpose of its production or creation	‘bread’ – ‘food’
origination place relation	object and its place of origin	‘bread’ – ‘bakery’
origination time relation	object and time of its production or creation	‘bread’ – ‘baking time’

In the EcoLexicon project, the product-instrument relation (*effected-by*) is “especially meaningful in those domains where human interaction plays an essential role, as is the case of environmental contexts” (Faber, León-Araúz and Prieto Velasco 2009, 10). Furthermore, the authors find the resultative relation (*result-of*) relevant when processes or entities are derived from other processes. This could also be treated as a temporal relation between successive phases (Section 4.2.6). In addition, Maroto and Alcina (2009, 236) add a state as a product of a process

(*process–state*). As in causal relations, a distinction could be made between objects, processes, and states as products. Purpose here is to be distinguished from the purpose of an action/activity (teleological relation) (Section 4.3) and from causal relations (Section 4.7).

The origination-place relation differs from the location relation (Section 4.2.2) since it involves concepts that refer to a product and the place or location where it is made or where it originates from. In contrast, the location relation is based on the connection between an object and the place where it is located. Similar to the origination-place relation is the relation that León-Araúz and Faber (2010) call *takes_place_in*. Their example “littoral drift *takes_place_in* the sea” refers to a natural phenomenon, which otherwise activates various types of causal relation.

4.5 Developmental relations

Developmental relations are based on objects that go through stages in various types of process (Table 6). The concepts in these relations may refer to stages of individuals (ontogenetic relation), species (phylogenetic relation), generations (genealogic relation), or materials (material development relation, e.g., ‘water’ – ‘ice’). An additional relation is the role-change relation based on the changing or developing roles of the same individual. Except for sequential connections, there are also simultaneous, co-occurring, co-influencing and co-existing phenomena, which activate even more complex concept relations and entire concept systems (Nuopponen 1994a, 101–103, 196–204; 2011, 10). In various fields, especially in Biology, there are systems and principles for creating this type of concept system.

Table 6. Developmental concept relations

Developmental concept relations	Based on relationships between	Examples
ontogenetic relation	development phases/ stages/ stages of individual	‘child’ – ‘adult’; ‘infancy’ – ‘adolescence’; ‘puberty’ – ‘adolescence’
phylogenetic relation	development stages of a species	‘grey wolf’ – ‘domesticated dog’
genealogic relation	generations	generations in a family tree
material development relation	development stages of a material	‘water’ – ‘ice’; ‘flour’ – ‘dough’ – ‘bread’
role change relation	changing roles of the same individual	‘presidential candidate’ – ‘president elect’ – ‘president’

4.6 Interactional relations

Interactional concept relations are based on the interplay between the objects of reference. Each type of interplay has its own relational properties and can activate networks of several concepts. Interaction is also included in many other relation types.

Within this group, transmission relations are based primarily on the relationship between a sender that sends something to a receiver. The classical communication model underlies this model, but it can be applied to many other fields and types of transmission (Nuopponen 1994a, 214–218). A transmission process may additionally involve one or more intermediaries, a path, and a tool (Table 7). Examples of this type of relation are sender-receiver, sender-intermediary, sender-object etc. (Nuopponen 2007, 207–209). Transmission is a type of activity and together with activity relations, they model most of the activity types.

Table 7. Interactional relations

Interactional relations	Based on relationships between:	Examples
transmission relation:	participants or objects in a process where something is transferred from one party to another	
– sender-receiver	– the ones who send and receive	concepts in communication process, business transaction, digital communication, etc.
– source-object	– sender/source of the object and the object	
– sender-intermediary	– sender and the mediator	
– intermediary-object	– mediator and the object to be send	
– intermediary-receiver	– mediator and receiver	
– tool-object	– tool used to transmit and the object	
– object-route	– object and the path	
– object-target	– object and the receiver	
representational relation	an object and its representation	‘senator’ – ‘constituency’; ‘term’ – ‘concept’; ‘place name’ – ‘place’
dependency relation	someone or something that is determined or conditioned by another and this other	‘employer’ – ‘employee’; ‘lender’ – ‘borrower’
correlation relation	two phenomena, when one changes, the other also changes	variables that vary or occur together; demand – price

The representational relation is based on the relation between an entity and another entity representing it (e.g., ‘concept’ – ‘term’, ‘people’ – ‘parliament’). Dependency relations are based on various types of economic, legal, and other connections that may exist when someone or something is determined or conditioned by another (e.g., ‘employer’ – ‘employee’), and the ones involved are mutually dependent on each other.

The example ‘adhesive’ – ‘adherent’ in Maroto and Alcina (2009, 236) for the relationship *concept-simultaneous concept (in space)* also exemplifies a type of dependency. The correlation relation is a reciprocal one between two entities, which means that when one changes, the other also changes (e.g., ‘price’ – ‘demand’).

4.7 Causal relations

As mentioned in the previous sections, the activity, origination, developmental, and interactional relations include causal components whereas the causal relation includes temporal components. As Marshman, L’Homme and Surtees (2008b, 145) point out, “the close interconnection of this relation with others, such as temporal succession and association, often poses challenges for its evaluation”.

Causality is a complex phenomenon, and in addition to ‘cause’ and ‘effect’, various other types of concepts are involved (Table 8). These refer to different types of cause (agent, action/event, circumstances, etc.), effect (resulting events, states, products) as well as patients, symptoms, consequences and countermeasures and cures. The concept system becomes even more complex when co-ordinate, parallel, or alternative concepts (e.g., alternative treatments for a disease) are taken into account (Nuopponen 2008).

Table 8. Causal relations

Causal concept relations	Based on relationships between	Examples
causator relations (agentive causal relation): causal agent – resulting event, state, product	causing agent and the effect	‘SARS-CoV’ – ‘COVID-19’
causation relations (producing causal relation): producing cause – resulting event, state, product	causing action and the effect	‘SARS-CoV infection’ – ‘COVID-19’
explanatory causal relations (circumstantial relation): explanatory cause – resulting event, state, product	circumstances needed and the effect	‘exposure to SARS-CoV’ – ‘COVID-19’
symptom relations: effect – symptom, symptom – cause	effect and its symptoms; symptom and its causes	‘COVID-19’ – ‘loss of smell’
counteraction relations: counteraction – cause, effect, consequence, symptom, patient	preventing or curing measures and causes, effects, consequences, symptoms or patients they are addressing	‘vaccination’ – ‘COVID-19’
patient relations: patient – cause, effect, consequence, symptom	affected and cause, effect, consequence, or symptoms for it	‘respiratory tract’ – ‘COVID-19’
consequence relations: cause/ effect – consequence	cause or effect and consequences they are leading to	‘COVID-19’ – ‘reduced lung function’

Causal concept relations have been addressed in terminological research more extensively for instance by Marshman (2002), Marshman et al. (2008a, 2008b), Nuopponen (1994b, 2008), and Pasanen (2019).

5. Conclusion

In this chapter, concept relations were characterized, based on a selection of typologies. After defining the relevance of conceptual relations in Terminology, their meaning for terminology work and terminological resources were discussed. The last part of the chapter focused on four typologies of conceptual relations. Nuopponen (1994a, 2005), which is the most extensive, was used as a basis for comparison. The inventories of conceptual relations in OntoQuery, Maroto and Alcina (2009), and EcoLexicon provided definitions, additions, and modifications as well as possible new perspectives and directions for future development.

Since all of these typologies vary from one publication to another, not all of the relation types in them could be covered in this chapter. Nuopponen's typology is the result of a theoretical study whose goal is to create a general concept system for conceptual relations. Accordingly, it is a top-down process, which takes existing typologies and relation descriptions, including Wüster's, as the point of departure. It has also been tested in various fields. The other typologies of conceptual relations mentioned in this chapter are from multidisciplinary research projects, which have a specific goal and which pertain to a certain specialized domain. Consequently, the sets of relation types were finally narrowed down to those that were relevant to their project. With the help of these domain-specific studies, we were able to pinpoint some lacunae as well as the need for the further division of certain relation.

Our review shows that although typologies of conceptual relations differ, they also have many features in common. The distinctions can be explained by different approaches and disciplines, which not only contribute their own terms and concepts but also their own concept systems. Interdisciplinary terminology in this case comes from semantics, ontology and corpus studies. When definitions and examples are scrutinized beyond the superficial terminological differences (e.g., *concept relation* vs. *semantic relation*), few differences can be truly observed. Because of the different concept systems, however, inventories of conceptual relations are not fully identical though there are many similarities. Shared characteristics can be found at the lowest level of abstraction, whereas the greatest differences reside in the way that the relations are grouped and in how the top-level classification is structured.

The top-level division in the typology by Nuopponen is based on generic (syn. logical) and ontological relations. Accordingly, ontological relations are divided

into contiguity (in space or time), activity, origination, developmental, interactional, and causal relations. In addition, instrumental and role relations comprise groups of their own. The most common relation types, such as generic, partitive, temporal and causal, are well described and scrutinized in research reports and Terminology handbooks. Nevertheless, many of the previously mentioned relation types need to be looked into closer and studied in future research projects and method development. A step in this direction is the international ISO 704 standard, which is in the process of expanding its set of relation types (see chapter by Wright in this volume).

Concept relations have been regarded as important elements in the theory and methods of Terminology since the very beginning. In many ways, they are needed now more than ever, because concept relation information is crucial not only for terminology work but also for other purposes of knowledge transfer. There is a need for terminological and conceptual clarity in regard to specialized knowledge transfer. There is also a growing body of research focused on how concept relations are expressed in texts since an infinite amount of digital text is available. To access and organize specialized knowledge, ontology work has foregrounded concept relations. The development of various types of query system emphasizes the need to analyze context and syntax with a view to finding related concepts. For example, Artificial Intelligence systems are finding and producing a lot of data. The analysis of these data has highlighted the need to classify and organize it. In this type of endeavor, terminology, concept relations and field-specific conceptual systems are extremely relevant.

CHAPTER 4

Terminology and standards

Sue Ellen Wright
Kent State University

This chapter views terminology-oriented activities related to standards, focusing on communities of practice. It ranges from word-oriented models to concept-oriented terminological resources, then spans the full semantic spectrum. It addresses lexicography and discourse-oriented terminology management as well as taxonomies, metadata systems for Linked Open Data strategies, concept registries, and 'thing'-oriented repositories. It cites terminology orientation, i.e., concept orientation, term-like units, semantic ordering systems, and harmonization aimed at a Single Source of Truth. ISO standards address discourse-oriented terminology management, including translation-oriented terminology management and team-oriented industrial environments, terminology exchange, and interoperability. Tight coupling in constrained resources and effects on relative semantic strength with respect to openness and machine accessibility round out the high end of the semantic spectrum.

Keywords: terminology standards, semantic spectrum, linked open data, ontologies, information systems, thing-oriented resources, tight-, loose-, weak-, strong semantics

1. Introduction

This chapter discusses and profiles terminology-related standards across a spectrum of language resources, focusing specifically on concept-based terminology work, ranging from discourse-oriented terminology resources used in the production and translation of human-readable texts, and extending on to terminological applications in varieties of knowledge-based, machine-processable resources. In so doing, it spans the evolution of terminology standardization. Even before World War II, but particularly in the period after the war, terminology standardization began in the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), when standardizers realized they had to clearly define critical terms in technical standards. With the digitization of text, the

disappearance of paper, and the growing ability to process virtually any text or data resource automatically, coupled with web-based markup and processing of semantic content, the 21st century ushered in the era of machine-sensible, ontology-oriented terminology resources that are no longer confined to human readability, but rather designed to organize knowledge in machine-parsable ways. The spectrum traverses the theoretical range of this volume, ranging from the General Theory of Terminology (see also Humbley this volume; Candel this volume) to thesaurus management and knowledge organization systems (see also Meyer in this volume), and ontologies (Montiel-Ponsoda this volume), but also touches on Corpus-based Terminology (Marshman this volume; Bertels this volume).

The various standards addressed here are linked to their relevant communities of practice (CoPs). In many cases, their associated standards bodies have divorced themselves from full form names, having terminologized their original acronyms and initialisms, resulting in ostensibly enigmatic, but widely used, short forms, whereby ASTM Intl. (International) was once (but is no longer) the American Society for Testing and Materials, or SAE was originally known as the Society of Automotive Engineers. This is a world of alphabet soup, where short forms abound because no one ever spells them out. An effort will be made to explain abbreviated forms on first usage, but a bookmark inserted in the Abbreviations page may be highly recommended. Abbreviated forms that are not specific to the topic at hand will be explained in footnotes.

Concept-oriented terminology management notwithstanding, this approach also touches on lexicographical and more encyclopedic resources, recognizing the role that terminology plays in related resources. Terminology resources are considered in all their forms as a continuum of potentially processable content, which reflects the trending view across the field.

Figure 1 breaks down language resources into different types or classes. All resources considered here either contain, document, or index language-related content. Some will be treated with more detail than others because they manifest different degrees of adherence to a terminological core. The abbreviated forms in the graphs will be discussed in context.

The definition of *terminology resources* taken from the ISO Technical Committee 37 *Vocabulary* standard ISO 1087:2019 itself poses a problem:

terminology resource
 terminological data collection
 TDC
 collection of terminological entries

Note 1 to entry: Terminology resources may be in paper or electronic format, e. g. paper dictionaries or glossaries, CDs, DVDs, databases or termbanks.

(ISO 1087:2019, 3.7.1)

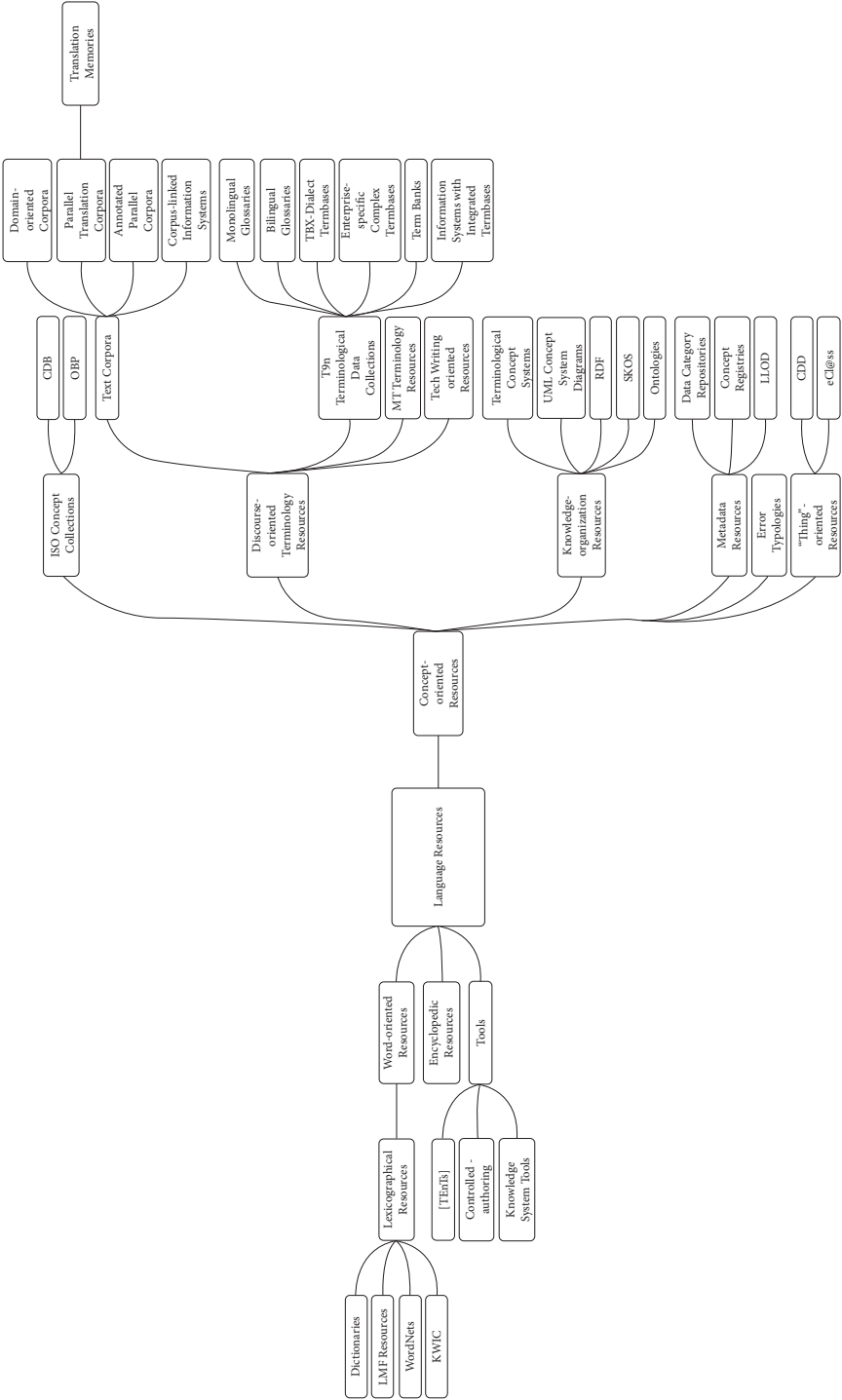


Figure 1. Standardization and language resources

The definition limits the concept in question to formal concept-oriented entries as treated in discourse-oriented environments, whereas this chapter also provides a view of extensions into other areas of the semantic spectrum. Indeed, some working groups in ISO Technical Committee 37 (TC 37) for *Language and Terminology* have split this definition with the observation that *terminology resource* is a broader term than *terminological data collection*. Figure 1 further divides language resources into four primary groups: word-oriented resources, including encyclopedic resources, and tools, which are grouped on the left side of the diagram, and concept-oriented resources, which are the primary focus of this paper, on the right.

Figure 2 recasts at least some of the items included in Figure 1 viewed from the perspective of the common semantic spectrum, whereby various types of resource (depending on the focus of any given author) are arranged along a cline, with weak resources on the lower left and strong ones on the upper right. McGuinness (2003) attributes the original notion of the spectrum to a panel discussion at the Association for the Advancement of Artificial Intelligence (AAAI) held in 1999, but Almeida, Souza and Fonseca (2011, 10) credit Daconta and Obrst for popularizing the basic model of the spectrum (Daconta, Obrst and Smith 2005; Obrst 2006). In this view, the characterizations of ‘weak’ and ‘strong’ reflect the degree of the explicit expressivity inherent in a resource, where strength is manifested by machine-processable semantic information, which at its most powerful appears as extensive relations documented in logical metadata. As will be demonstrated by several examples, there is a trend for some resources positioned low on the cline to adopt metadata solutions that move them up the cline in terms of accessibility

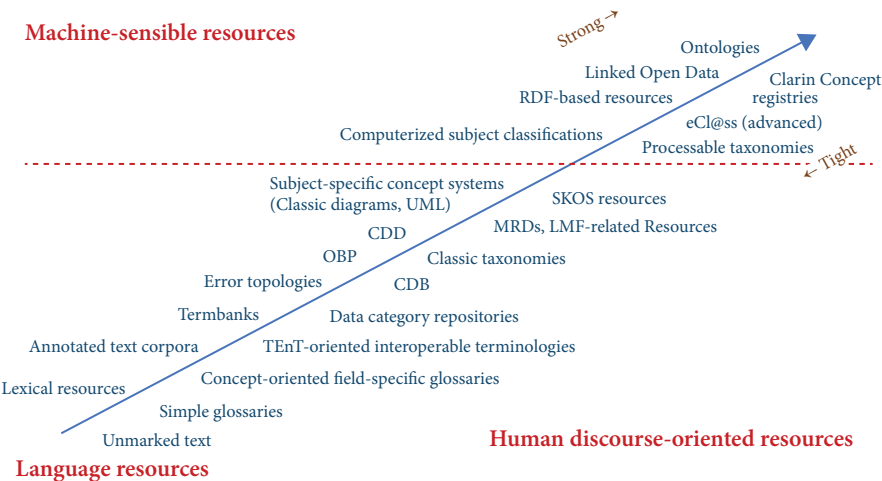


Figure 2. Language resources from Figure 1 projected on a semantic spectrum

and processability. By the same token, consolidation behind various kinds of domain-specific firewalls consigns otherwise promising resources to their position as weak or tight options.

Although many representations of the spectrum focus on the upper levels of the cline, the novel arrangement shown in Figure 1 plots standards-related resources representing all types of information products, beginning with weak, unmarked texts intended for humans to read, and progressing to rigorously logical, metadata-rich instruments designed for machine readability on the Semantic Web. Viewed from this perspective, the weakest resource would be a text or word list printed on paper, and the strongest would be a fully populated ontology.

A further distinction introduced by Obrst (2006) is between tight and loose (tightly coupled and loosely coupled) resources, whereby tight resources are rigorously bound to a single application, subject field, or computing environment, perhaps even locked up behind a pay-to-play firewall, and are not available or as available for open access and machine readability as are other resources appearing higher along the spectrum (Sandoval 2020).

2. Word-oriented and lexicographical resources

First of all, lexicographical dictionaries and related resources focus on words and their potentially multiple definitions, which is to say, words and their one or more meanings. In contrast, entries in terminology resources treat individual concepts in single entries, designated by one or more terms and potential equivalents in other languages if they exist in the universe of discourse. Word-oriented resources comprise works such as lexicographical dictionaries, more random glossaries, and word lists, which are nonetheless weakly terminological because they can be processed simply as text, and they provide a certain level of terminological information. Encyclopedic resources comprise virtually all text that is not configured as either lexicographical or terminological in nature, hence the bulk of all written text. These resources also contain terms, and can in many instances be incorporated into cohesive, subject-specific corpora, which take on a stronger terminological nature when they are filtered or sorted for subject matter specificity, and particularly, if they are marked up with appropriate metadata, so that they can be processed using corpus management tools.

Tools are placed here on the lexical side of Figure 1 as a matter of convenience. They can be applied at virtually any level of the graph. The division into Translation Environment Tools (TEntTs, Zetsche 2008), controlled authoring, and knowledge system tools is an over-simplification, but this also reflects the fact that tools themselves are not the primary subject of discussion. No effort has been made

to document all tools and processing environments. Keyword in Context (KWIC) and WordNets, which provide a range of semantic relations on the lexical side of the spectrum, are cited as examples of approaches in this area.

The apparent distinctions in Figure 1 are not necessarily as clean as they might appear at first glance, and representations of the general field in question vary with different perspectives and approaches. The inclusion of subject-field-specific text corpora, although they are shown here in conjunction with concept-oriented varieties, can themselves be used to compile lexicographical resources as well as terminological ones. In this regard, the “text corpora” component of the diagram could be replicated at several nodes of the diagram if these other segments were to be expanded to place additional focus on the textual origins and context of words and terms. The placement of text corpora should then be viewed as exemplary of the role that they may play in a variety of environments.

3. Terminology resources

Terminology resources must meet the following criteria:

1. Concept orientation: the language resource is based on the identification of the notions, the ideas underlying thought and communication. Concepts are represented by term-like elements, and delimited by clear definitions stating the characteristics that comprise them and distinguish them from other similar elements in their class.
2. Term-like elements: they may be words, compounds, phrasal strings, or identifiers that serve as normalized or even standardized identifiers or designations to represent individual concepts. In many systems, concepts or concept entries are represented by both terms and other human-readable designations, as well as in parallel non-mnemonic identifiers. Terminology elements can be examined for their degree of termhood (Kageura 1996; ISO 26162-3, under development, Clause 4).
3. Concept ordering: Term units and concept entries are frequently organized in ordering systems reflecting hierarchical (generic or partitive) or associative relations. These concept systems may in some cases be document-oriented and organized as isolated concept groups, as is often the case for terminology collections created for individual documents or standards. Terminology resources created for short projects and individual or small sets of documents are typically referred to as ‘ad hoc terminologies’ (*terminologie ponctuelle*, Nakos-Aupetit 1977). In other cases, an effort is made to create thematic terminologies and systematic ontologies for extended bodies of knowledge, expanding to comprehensive coverage of entire subject domains.

4. Single Source of Truth: communities of practice function as harmonizing or even normative bodies to establish a ‘Single Source of Truth’ (SSoT), also known as ‘single point of truth’ (SPoT), which is to say that concepts making up the resource are collected in a single resource for purposes of definition and specification. Even if the resulting collection may be duplicated elsewhere, all related content should be generated by a direct link to or controlled export from the SSoT. This criterion is designed to prevent issues involving version control or conflicting definitions, explanations, or ordering principles (Dragon1 2021).

4. Communities of practice and standards efforts

4.1 Lexicographers and lexicographical resources

As noted, the lexical and lexicographical resources shown on the left side of Figure 1 are fundamentally word-oriented, but nonetheless in many cases facilitate extension into more terminological environments. One major standardization effort is associated with the development of LMF, the Lexical Markup Framework, in ISO/TC 37, SC 4. LMF Core is word-oriented, as reflected in the definition of a lexical entry. It should be noted that the reference to ‘container’ here reflects the fact that this standard is UML-oriented (Unified Modeling Language, OMG 2017), which graphs classes of related objects in boxes called containers. In this definition, the ‘form’ referenced is a word form, usually a lemma.

lexical entry

container for managing one form or several forms and possibly one or several meanings in order to describe a lexeme

(where “form” is any given instance of a lexical unit; ISO 24613-1:2019, 3.6, 3.2)

The LMF core model provides for the documentation of typical lexicographical elements: grammar, word forms, multiple senses, definitions, and external references. Additional parts of the standard provide extensions in the form of an Etymological extension, description of Machine Readable Dictionaries (MRDs), a TEI (Text Encoding Initiative) serialization, and the Lexical Base eXchange (LBX), which provides an XML¹ serialization. These supplemental parts of the standard support the interaction of machine readable dictionaries (MRDs) with other resources to enable natural language processing in more complex environments.

1. XML, the World Wide Web Consortium (W3C)’s eXtensible Markup Language, is a framework tagging language that can be freely configured according to set rules in order to display defined data sets on the WWW. It is considered to be both human and machine readable (W3C-schools, Introduction to XML).

For instance, the OntoLex-Lemon (Ontology Lexicon Community Group + Lexicon Model for Ontologies 2016) has created an RDF² format that facilitates the publication of lexical data expressed in LMF that negotiates a linkage to the Linguistic Linked Open Data (LLOD) environment and Natural Language Processing (NLP) applications (see also Montiel-Ponsoda this volume). Thus, MRDs are not themselves ontologies, but via LMF, they may contribute valuable linguistic information that functions in concert with ontological resources. In this respect MRDs are stronger along the cline presented in Figure 2 than standard human readable dictionaries, but do not yet cross the line into the level of complete knowledge organization resources.

4.2 Terminological language resources (ISO/TC37/SC1, 2, and 3, discourse oriented terminologies)

Terminology management in the area of standards development began initially with what is categorized here as ‘discourse-oriented terminology management’, which is intended to provide standardized or organization-specified terminology for use in both monolingual writing and bi- or multilingual translation and localization environments. As a community of practice, the creators of these resources tend to be translators, technical writers, and in some cases, librarians. They also create visually-oriented concept systems for human examination, which in the past at least have not been machine-processable.

Librarians traditionally have concerned themselves with subject field classification systems and knowledge organization systems associated with book curation and research query environments. Under the auspices of the W3C (Worldwide Web Consortium) Semantic Web Deployment Working Group and its efforts during the first decade of the 21st century to create a set of RDF-based data representation resources, SKOS (Simple Knowledge Organization System) has been developed for the purpose of representing thesauri, classification schemes, subject heading systems, and taxonomies in the context of the Semantic Web. The critical distinction between SKOS-related information and more powerful ontologies is that SKOS is generally focused on hierarchical systems and associative networks, whereas more powerful ontologies provide functionalities for expressing logical axioms (W3C 2009). Nevertheless, SKOS-based resources form a base for development in the direction of linked open data.

2. RDF, the W3C’s Resource Description Framework, is a machine-readable (but not human-readable) XML formalism designed to describe resources on the web (W3Cschoools, XML RDF).

Subject-specific taxonomies are not discussed here in detail, but it should be noted that they are among the oldest classification systems, with multiple approaches in some disciplines (Biology, for instance), and strict control in others (CAS, formerly referred to as the Chemical Abstracts System, which rigorously controls the assignment of various levels of chemical nomenclature).

The wide range of knowledge-based, machine-processable systems are the province of subject-area specialists in a wide variety of fields, coupled with a focus on resource metadata and Computer Science solutions. Aside from Web functionality, these applications are particularly useful in commerce and industry, as reflected in 'thing'-oriented resources that can be used to coordinate the exchange of products, processes, and services in industry. Here the stakeholders range from manufacturers, logistics experts, and sales personnel to end users.

5. Standardization linked to CoPs

5.1 Human-readable terminology resources

Standards-related and discourse-oriented terminology practice is the primary subject arena of ISO Technical Committee (TC) 37/SC 1, Language and Terminology, and the stated topic of its core standard ISO 704, *Terminology work – Principles and methods*. This standard, along with most of the terminology-related standards produced in the committee, is concerned with defining and organizing concepts associated with terms used in human-comprehensible discourse and human-readable texts.

Terminology work according to this document is concerned with terminology used for unambiguous communication in natural language, in particular special languages. The goal of terminology work as described in this document is, thus, a clarification and standardization of terminology for communication between humans. Terminology work can be used as input for knowledge modelling, information modelling, data modelling and classification, but this document does not cover the relation with these fields. (ISO 704)

ISO 704 spells out the basic principles of object-concept-term relationships, reflecting linguistic theory that has developed over the past century, drawing on principles that are for the most part familiar to students of formal Linguistics, Translation Studies, Technical Communication, and similar disciplines. The standard has been reviewed and revised through the decades, becoming more complex and more oriented towards the discipline of formal Linguistics.

ISO 704 also introduces the reader to the definition of concepts using critical characteristics, which in turn enables the creation of visual concept systems for the logical ordering of terminological hierarchies and associated relations. Until the late 20th century, traditional subject classification and taxonomy systems, along with visual (non-machine-readable) graphs served to represent concept relations. The most recent edition of 704 also shows examples using UML modeling to represent concept systems as specified in ISO 24156-1, *Graphic notations for concept modelling in terminology work and its relationship with UML*. Another aspect of the current revision of ISO 704 reflects a refinement of types of associative relations.

The Unified Modeling Language (UML) defines a metamodel and an abstract syntax used by system architects and software developers for representing modular structures underlying database and formal language designs. It provides tools for implementing software designs or representing other logical structures in business and management. It should be noted, however, that even moving resources like LMF and ISO 704 in the direction of UML still reflects relatively weak ontological practice, as reflected in Figure 2. This statement does not, however, represent any kind of negative reflection on the power of the system to perform the task for which it was designed.

ISO 704, ISO 1087 *Vocabulary*, and TC 37/SC 2's ISO 10241-1:2011/2012 *Terminological entries in standards* comprise horizontal standards that provide requirements for the creation of terminology standards and the terminology entries in ISO and IEC standards. There are also two major new terminology standards in process: ISO 20539 for *Translation, interpreting and related technology*, and ISO 24183 for *Technical Communication*. Although an initial version of 20539 has already been completed, work continues on terminology associated with leading edge developments in the rapidly evolving interpretation field. Technical communicators from a number of IEC working groups have come together under the umbrella of TC 37/WG 10 to compile a comprehensive collection of vocabulary that can serve as a truly horizontal standard for their many standards involving technical writing environments.

5.2 ISO concept collections

The standards cited in the previous section govern the first major resource cited in the *Concept-oriented Resources* group: *ISO Concept Collections*.

As early as 2009, ISO initiated an ambitious project dubbed the *ISO Concept Database* (CDB), which was originally designed to incorporate terms and definitions from ISO standards, graphical symbols, codes (language, country, currency, etc.), units of measurement, product properties, data dictionaries of various types, and (unstated) other resources. The plan was to provide a workflow for individual

concept standardization similar to the workflow used for the elaboration of ISO documents. Information published at the time indicated that the resource would be a pay-to-view subscription service (InfoTerm 2009), indicating a trend on behalf of the standards bodies to seek to monetize their language resources beyond charging for the standards themselves, which would have positioned the information behind a paywall where it would not be available to open searching on the Web.

Today a click on the old CDB link will resolve to the *ISO Online Browsing Platform* (OBP), reflecting the fact that the overall plan for the CDB underwent a significant rework, to be relaunched as the OBP in 2012 (FIT 2012). The current resource provides human-reader access to individual terms and definitions extracted directly from all ISO standards, plus the Introduction, Scope, complete Terminology Section, and References. Materials can be accessed in English, French, Russian, and Spanish. The original announcement indicated intentions to provide XML output, which would point at least to portability, and possibly machine readability. Nevertheless, it is not apparent that this functionality has ever been implemented.

The fact that the OBP, unlike the originally planned CDB, is available free of charge to users on the web runs counter to the general practice of ISO and IEC and related standards bodies. ASTM (formerly the American Society for Testing and Materials) maintains its *DEFONLINE, Online Terminology Dictionary of Engineering Science and Technology*, but charges a yearly subscription rate. It would be interesting to determine whether providing a good bit of free information translates into increased sales and public awareness. In contrast, W3C and related standards are available free of charge open to download or accessible for machine processing. A fundamental business model that locks a resource behind a pay-to-play, organization-specific firewall relegates that resource to a tight position that prevents it from making a significant contribution to any goal for accessibility and machine reasoning on the global Web.

ISO country codes (ISO 3166) are available in the OBP, and along with a complete directory of ISO symbols. The original CDB intention of including database content apart from other print standards documents has not been realized. The system is not machine sensible, nor can the data or subsets thereof be output in any parsable (e.g., XML, RDF, or the like) form. This overall lack of processability places the CDB below the dividing line in Figure 2, separating human-readable from machine-processable resources.

The lack of language codes (not strictly a terminology resource) is evidence of the difficulties involved in seamlessly representing data in one system that is generated and normally resides elsewhere without violating the Single Source of Truth (SSoT) principle. ISO 639-2 three-letter language codes are an integral part of the US Library of Congress (LoC) MARC (Machine Readable Cataloguing) record

and hence reside in their original form on the LoC website, and the SIL Ethnologue 639-3 three letter codes are housed on the substantial SIL website, along with a wealth of additional language-related information. While it would be extremely useful to access the language codes in the same reference as the other OBP resources, a solution needs to enable parallel representation of these data without introducing inconsistencies.

5.3 Specialized TC 37 standards

5.3.1 *Theoretical principles and pragmatic practices*

TC 37/SC 2's ISO 12616-1, *Fundamentals of translation-oriented terminology* (T9n) describes the significance of terminology management in the broader context of translation processes, especially with respect to complex translation workflows within potentially extensive multilingual work environments. It defines the roles played not only by translators and terminologists, but also of the so-called terminology worker and the subject matter expert (SME) in creating, maintaining, and using terminology management systems (TMS), as opposed to and in conjunction with the creation of translation memories (TM). It clarifies basic principles, such as distinctions between prescriptive and descriptive approaches, as well as between systematic and ad hoc terminology management.

A major difference between this standard and those that introduce general principles is its concentration on process and team building in enterprise environments. The approach to creating concept entries is more pragmatic perhaps than in ISO 704, with an emphasis on efficiency and effectiveness in organization workflow. It outlines the relationship between terminology resources and other components of Translation Environment Tools (TEntTs). It includes a brief reference to the TBX *Termbase eXchange* standard, proposes a data model that for all intents and purposes comprises a TBX dialect (see discussion of ISO 30042 below), and explores skills and competences associated with different aspects of the terminology management process.

TC 37/SC 3, ISO 26162 *Management of Terminology resources*, has elaborated a more comprehensive set of pragmatic standards, *Terminology databases, Part 1: Design, Part 2: Software, and Part 3: Content*. Part 1 outlines the terminological metamodel specified in ISO 16642, the *Terminological Markup Framework* (TMF), and relates the respective data categories from datcatinfo.net to their respective positions in the metamodel. It outlines their properties and discusses concept relations that are typically recorded in terminology databases (termbases), articulating the basic principles of dependency, repeatability, granularity, elementarity, and data modeling variation. Part 2 aggregates the various standards already cited

above in order to delineate their application in the creation of a database definition that complies with the TMF metamodel. It outlines requirements for terminology management systems in order to ensure proper functionality in typical terminology management workflows, with special emphasis on search functionalities, filtering, data representation, and exchange functionalities.

Part 3 focuses on ensuring the quality of termbase content. Significant topics include criteria for identifying and including concepts and related terms to be covered in the resource, establishing the concepts of ‘unithood’ and ‘termhood’ (as opposed to words in general language). Strong emphasis is placed on the extraction of terms from relevant corpora based on an organization’s specifications with respect to purpose, correctness, and usability. In addition to translators and terminologists, this set of standards addresses a wider net of stakeholders with respect to user roles, including technical writers, subject matter specialists, and others. Special emphasis is placed on the quality analysis of individual concept entries as a function of terminology management system (TMS) audits within the overall enterprise quality assurance framework, culminating in a proposed model for a sample error typology to be used together with an evaluation model. Later this chapter will take up error typologies as a separate terminology resource in their own right.

5.3.2 *Interchangeability and interoperability*

TC 37/SC 3 is also responsible for the ISO 30042, *TermBase eXchange (TBX)* standard, and ISO 12620, *Data categories* (which has segued into the two-part 12620-1, *Specifications*, and Part 2, *Repositories*.) A termbase is a terminology database consisting of a collection of concept-oriented entries. A terminology entry might look something like the concept entry for the term *data category* that follows, where for this representation, the data category names appear in bold face.

Term: data category

Definition: class of data items that are closely related from a formal or semantic point of view

Example: /part of speech/, /subject field/, /definition/.

Note: A data category can be viewed as a generalization of the notion of a field in a database. (Based on ISO 30042, 3.8)

ISO 30042 appeared in its most recent (version 3.0) incarnation in spring 2019, soon to be joined by related spin-off technical reports and specifications. The ISO 24633 multi-part series of Technical Reports provides guidance on applying various serializations, representations, and/or mappings for TBX (RNG, RDF, JSON, DCT style), and ISO Technical Specification 24634 incorporates concept relations and subject field information into the TBX environment. Historically, the TBX standard

evolved in parallel with ISO 12620. The original SGML³ (Standard Generalized Markup Language) termbase exchange standard (ISO 12200:1999) was based on two components: a data model, which came to form the basis for ISO 16642, coupled with a catalog of data categories comprising the data fields and any pre-defined values incorporated into concept-oriented term entries. For instance, */term/*, */definition/*, */example/*, and */note/* are all open data categories, in that they can take any text content that complies with the definition of that particular data category. The examples */part of speech/* and */subject field/* are usually configured as closed data categories, where values must be chosen from a picklist, for instance, */noun/*, */verb/*, */adjective/*. These typical picklist values are considered to be simple data categories since they represent values and cannot have sub-value children of their own.

It is assumed that in a given language, it is possible to standardize values for all termbases documenting a field like */part of speech/*, but */subject field/* will feature customized values for a particular environment. As noted above, these kinds of values will be accommodated according to ISO 24634, where subject field and concept relation values can be represented in SKOS in the backmatter of the TBX file, thus enhancing the machine processability of the termbase. This expanded direction for TBX reflects a trend in Terminology Management System (TMS) design to include more sophisticated representation of concepts systems. TermWeb^{*} TMS, for instance, includes the option to create linked graphs among related concept entries, and the Coreon^{*} TMS 'combines taxonomies with terminology to create and deploy Multilingual Knowledge Systems' with the goal of supporting search, machine learning, and interoperability with the Internet of Things (IoT). Actual instantiation of these functions in a given software application will vary depending on any particular system.

Collected from a wide variety of terminology resources and tools applications available to the growing field of computerized terminology management in the early 1990s, the resulting set of data category types and prescribed values was incorporated into the original ISO 12620 of 1999 (see also Wright and Budin 1994). This study was conducted in parallel with the initial elaboration of the TBX standard, which has undergone a series of transformations, first from the SGML serializations TEI-Term and MARTIF (ISO 12200), but then into the XML TBX 2.0 version of 2009. The 2019 TBX 3.0 variation was reconfigured to integrate with the OASIS/ISO TC 37/SC 5 ISO 21720:2017 XLIFF, *XML Localisation interchange file format*, as its terminology component. Although TBX retains its original data model structure, it had always proven to be problematic because its extreme flexibility made it difficult to exchange termbases that exhibit variations within that declared

3. SGML (Standard Generalized Markup Language), introduced as an ISO standard in 1989 (ISO 8879), was originally designed to provide flexible metadata markup of written documents, but gave rise to a whole family of markup languages, including XML, RDF, and RNG.

framework. TBX-Basic is an XML format reflecting one possible dialect of TBX, originally proposed by the now-defunct LISA/OSCAR⁴ working group. Initially configured to comply with TBX 2.0, the current version would involve some slight modification to comply with TBX 3.0. TBX-Basic was intended to reflect “industry best-practices for terminology databases that are used for content production purposes such as authoring and translation” (TerminOrgs). The premise behind this effort was to provide a uniform data model designed to ensure interchangeability and interoperability.

TBX 3.0 provides for declaring a TBX dialect as the specific data model for any given resource, which enables the identification of master files that can be used to ensure efficient and accurate interoperability or importation of data. Voluminous public resources posted at tbxinfo.net provide support information and utilities designed to facilitate the application of the standard. The ISO 24633 spinoffs cited above will enable the integration of the standard into existing termbase applications, most of which are designed to interact with Translation Environment Tools (TEnTs) as they adopt XLIFF 2.0. As noted, ISO 24634 is intended to accommodate subject field and concept ordering information in the TBX environment, thus moving TBX closer to the knowledge management trend in terminology resources by incorporating SKOS notation as a component of TBX backmatter in order to document lower-level knowledge-based information inside a TBX export file.

5.3.3 *The translation error typology*

A translation error typology is a systematic list of terms designating the kinds of errors that can occur in translated texts. Currently two projects are underway involving the analytic evaluation of translation quality: one in ASTM (ASTM WK46396, Analytic Evaluation of Translation Quality) and one in ISO (ISO 5060, Translation services – Evaluation of translation output). The two working groups have agreed on a uniform high level error typology to be used with statistical metrics to calculate normed scores that facilitate the comparison of evaluation results across variable factors. (The MQM 2021; W3C/MQM 2021; TheMQM 2022).

A second ASTM project, WK54884, *Holistic Quality Evaluation System for Translation*, focuses on overall quality aspects with less reliance on statistical methodologies. Although these documents comprise translation-oriented standards, the creation of error typologies itself is an exercise in terminology management and prescription, and the ASTM and ISO committees (which share some overlap in membership) have formed a joint task force to coordinate terminology between the two groups in order to ensure compatibility among the three projects.

4. LISA/OSCAR, Localization Industry Standards Association/Open Standards for Container/Content Allowing Re-use

Although the American Translators Association (ATA) Accreditation Committee began creating their *Framework for Standardized Error Marking* (2017) in the 1990s, and it has been widely adapted in academia as a model for evaluating student work (Doyle 2003; Koby 2015), the first industry-oriented standard metric for measuring the quality of translations dates from the SAE J2450 *Translation Quality Metric* in 2001. Originally configured with seven declared error types (*wrong term, syntactic errors, omissions, word structure and agreement, misspelling, punctuation, miscellaneous*), the standard provided for a score sheet for tallying errors and assigning weights (*Serious* and *Minor*), with the goal of producing normalized scores designed to reflect the quality of translations. The current version of the standard retains the seven major error types, and supplies sub-categories for more detailed evaluation. Criteria for assigning major error status reflect standard risk analysis practice in the automotive industry:

- Harm to the user, operator, patient, or consumer;
- Damage to the product, or equipment used with the product;
- Major misinterpretation of the author's intent, or the client's product or services;
- Damage to the reputation of the client. (EC Innovations 2019; Woyde 2001)

The current ASTM list (posted on the *theMQM* website) expands on this sort of typology, citing seven primary or 'top-level' dimensions: *Terminology*, *Accuracy* (other than Terminology), *Style*, *Linguistic conventions* (formerly *Fluency*), *Locale conventions*, *Audience appropriateness* (formerly *Verity*), and *Design and markup*, with a collection of sub-types assigned in each of these primary dimensions. This list has a long history, starting with the LISA (Localization Industry Standards Association) *QA (Quality Assurance) Model*, which is still available and continues to be used in some venues. It was replaced by the *TAUS/DFKI MQM model* (TAUS 2015; DFKI 2015), and is currently maintained by the MQM working group, which attempts to retain clear mappability. The ASTM revision is aimed at fleshing out the metric and providing detailed guidance for evaluation procedures and determination of root causes for error. Much of the work consists of carefully defining the items in the list and structuring the ordering system, which places this endeavor clearly in the range of terminology work.

Dalla-Zuanna and Kurz (2019) describe the evaluation process using these models. Errors are classified according to the predefined, weighted categories and points are tabulated in a metric score card, enabling the calculation of normed scores, which can then be measured against a predetermined threshold value stated as the desired goal for a particular text based on user expectations. This means, for instance, that different translations performed for different purposes may have different target values for the anticipated normed score. Their list of top-level error dimensions parallels the MQM list, although the current ordering for the *theMQM* website reflects considerable additional development since completion of the article.

Establishing a benchmark score for individual translations can contribute to the extrapolation of overall quality scores for a translation service or department, particularly in environments where Quality Assurance data are collected in an enterprise-wide ISO 9000 environment. Furthermore, the determination of error type for specific errors can contribute to two major purposes: the determination of root causes which may, based on this knowledge, be corrected in order to improve process; and the identification of the effects that any particular serious error may have based on the future use of the target text. All these models tend toward the exhaustive when expanded to include a full range of sub-types, but as indicated in Dalla-Zuanna and Kurz (2019), the intent of all the models cited here is to provide rich options for individuals to pare down the list and create customized score sheets designed to meet their specific quality evaluation needs.

6. Knowledge organization and metadata resources

6.1 Knowledge organization systems as terminology resources

Various types of knowledge organization systems frequently appear in integrated environments together with other modalities. Knowledge organization resources also rely on the identification and definition of concepts, together with the assignment of labels or terms to represent concepts. The critical factor setting true knowledge organization systems off from discourse-oriented terminologies is the explicit expression of concept relations, often through the use of metadata markup, not just for human interpretation, but in machine-sensible or machine-parsable structures that allow for the automatic processing of the resulting information structures. There are a wide variety of approaches developed by different communities of practice for their individual knowledge processing needs (Almeida, Souza and Fonseca 2011; Oberst 2006). Such resources are often used in tandem with terminology management systems, as cited with regard to Coreon⁵, or can stand alone and be applied in many different environments with a wide variety of functions.

As noted earlier, creators of discourse-oriented terminology resources tend to be writers and translators, and their concept systems tend to be visually oriented in their presentation of information. In contrast, the creators of ontologies and other high level resources more likely have backgrounds in Computer Science and Knowledge Engineering, and their (often RDF-based) serializations (SKOS, OWL,⁵

5. OWL, the *Web Ontology Language*, is a W3C-based family of XML-based languages designed for the logical, machine-processable representation of classes of objects and the relations among them (W3C 2012).

and other DOL⁶ -enabled systems) provide a variety of modalities for representing and processing knowledge organization systems in private closed environments or across the open Semantic Web. These systems appear above the red line in Figure 2, indicating increasing degrees of logical processability and openness, although the level assignments may vary in different representations of the spectrum.

Knowledge organization systems have arisen out of the special needs and mind sets of different CoPs, with the result that seemingly similar systems are incompatible or require special treatment or conversions in order to meet the needs of new applications as they arise, reflected in Kuhnian paradigmatic gaps. Despite the fact that language resources share the criteria outlined at the beginning of this chapter, and despite the fact that many of the same features recur in many systems (words/terms/identifiers/designations of some sort, grammatical information, definitions), past efforts to reuse the same elements across systems have proven problematic, if not futile. The reasons for these discontinuities lie in the significant differences in applications and environments. Many full ontologies can be rendered interoperable through the agency of DOL, provided that their models and specifications are appropriately identified, but particularly in the case of resources that are tightly associated with very specific applications or locked behind firewalls, sharing resources across CoP boundaries may be problematic or impossible.

6.2 Terminology theory and communities of practice

The ISO requirement that all ISO and IEC standards use ISO 704 as their guide for creating terminological entries in standards meets with frustration on the part of individuals who are not necessarily trained in modern linguistic theory or who want to go well beyond the limits of this particular standard. Consternated programmers and engineers often find it too ‘philosophical’ (where linguists simply see it as based on the fundamental principles of their discipline) and critique it for not adequately addressing their needs with respect to terminological aspects of knowledge management systems and ontologies. Although the 2021 draft standard clearly states that this area is outside its scope, this insight has not been apparent to many current potential users. ISO JTC, SC 32 has recently addressed these concerns by launching work on a new standard, currently titled: ISO/IEC AWI 5394, *Information technology – Criteria for concept systems*.

6. DOL™, the *Distributed Ontology, Model, and Specification Language*™ provides interoperability among classification and ontological systems by checking for coherence among models and specifications (OMG 2018).

The standard is still in its initial stages, so it would be premature to try to analyze whether the explanation of concept theory presented early in the discussion will be more comprehensible for this audience than the presentation in ISO 704. The important distinguishing criterion is that this document specifically extends the spectrum of the topic out into the ontological end of the cline in Figure 2 to include systems that implement “automated reasoning based on certain rules” together with an emphasis on information organization and the navigation of data resources (ISO AWI 5394, Section 5). It clearly spells out criteria for effective concept systems and outlines an abstract framework for representing a variety of relations within a concept system.

The draft standard proposes four major levels in its own version of the spectrum (called in the current version a ‘level structure’):

1. Weak structures, such as alphabetical glossaries, which are not even normally identified as concept systems per se;
2. Hierarchical concept systems such as the generic and partitive systems described in ISO 704;
3. Hybrid concept systems, including taxonomies, thesauri, and subject heading systems;
4. True knowledge-based concept systems and ontologies.

The bulk of the standard provides descriptions and examples of these levels and the many different approaches. It can be hoped that when the document is finished will provide an effective overview of the various systems that are currently in play.

6.3 Data categories and Linguistic Linked Open Data

Closely related to the knowledge organization systems cited here are other metadata resources, including specifically TC 37’s Data Category Repository (datcatinfo.net) and concept registries associated with the Linguistic Linked Open Data (LLOD) initiative. LLOD is made up of a consortium of collaborating researchers working with a variety of metadata resources to create a cloud-based network of open hyper-text links for a wide range of language resources, including corpora, lexicons and dictionaries, terminologies, thesauri, knowledge bases, linguistic resource metadata, linguistic data categories, and typological databases (McCrae and Cimiano 2018; Cimiano et al. 2020). Although some aspects of the linked data movement are not specifically terminological, the specification of related vocabularies associated with persistent identifiers (PIDs, ISO 24619) positions these efforts somewhere between thesauri and ontologies in many respects.

The temptation to merge systems despite fundamental modeling differences is reflected clearly in the history of the Data Category Repository. As noted above, the collection originated with the development of TBX and was originally published as ISO 12620:1999, which actually listed the content of the individual data category specifications printed on paper. Between 2008 and 2015, concerted efforts were made to incorporate the TC 37 *Data Category (Registry) Repository* with a project to document data elements from a wide range of language resources including terminology, linguistic metadata, morphosyntax, and other linguistic sub-domains that have mostly become the province of TC 37/SC 4 standards. Combining this sometimes disparate group of elements was based on the premise that they were all terminological in nature and all fit into the core pattern of ISO/IEC 11179 standards for the specification of metadata elements. It should also be noted that when ISO 12620:1999 was withdrawn, the standardized status of the original data category specifications dissolved as they took up residence in the ISOcat registry, with the intent that they would gradually be standardized one by one in the course of registry maintenance and administration.

The data model for the individual specifications as defined in the revised version of ISO 12620:2009 was fairly complex, consisting of three separate sections: (i) an administrative section designed for documenting standardization-related actions; (ii) a primary name and definition section not unlike a terminological entry; and (iii) a technical section which provided mechanisms for distinguishing open and closed data categories and specifying simple data categories associated with closed data categories. The standard also outlined the model and procedures for developing a 'standard as database' in line with the then current plans for the ISO Concept Database (CDB). Aside from attempting to meet the needs of multiple communities of practice, the design reflected special requirements imposed by ISO Central Secretariat in conjunction with their intention at the time for this Data Category Registry to become a standard as database under the aegis of the CDB. As recounted above, it was at this point that plans for the CDB dissolved, and at least the CDB-compliant segment of 12620:2009 became obsolete.

By 2015, a number of other problems had emerged. First of all, efforts to ballot individual data category specifications in line with the ISO guidelines for the CDB failed miserably. The notion of enlisting a significant number of individual TC 37 members to analyze and ballot on each single data category specification through a complex, three-stage process simply did not attract enough interest to make the process feasible. At the same time, it became evident that the various groups participating in the ISOcat project had very different goals and needs. Further changes in the orientation of the Max Planck Institute, which had served as an ISO Registration Authority (RA), accelerated changes in the organization of the revised resources.

This turn of events led to the decision to abandon ISOcat.org as an ISO registry and to divert the collection into two separate streams: a CLARIN-sponsored Concept Registry (CCR) and a TC 37/SC 3-defined Data Category Registry (later Repository, commonly referenced as the DCR). The former is organized under the auspices of the Meertens Institute and forms the basis for the semantic interoperability layer of the European CLARIN projects, which essentially means that this resource is designed for participation in the Linguistic Linked Open Data environment. The individual specifications are quite simple, consisting of a */Label/*, */Definition/* and a */Status/* value. (CLARIN 2015-present).

For its part, TC 37/SC 3 ported over 3000 selected specifications, including the entirety of the *Terminology* profile, from the original ISOcat collection to a carefully defined data model residing in the interverbumTech online TermWeb^{*} terminology data management system. The flexible, TBX-compliant data modeling capability of this resource enables compliance with data category specification structures, with the exception of critical links between closed and specified simple data category specifications, but the designers of the new system have used the built-in relations capability of the system to replicate those links. Hopefully in future, it will be possible, perhaps by building TBX 3.0 and compliance with the future ISO TS 24634 into TermWeb^{*}, to enable the import and export of this critical information along with other specification data using TBX import-export features. A further desideratum is to finally implement the original PIDs associated with each specification as live links in TermWeb^{*}, which fulfills one of the requirements of the Linked Open Data environment for open persistent Uniform Resource Identifiers (URIs). (The information is in the current specifications, but they have not been activated to provide open navigation capability.)

Although the current standards bearing the ISO 12620 number only prescribe the formulation of the specifications themselves and the administration of data category repositories, the actual content of the specifications technically resides outside the umbrella of standardization. This approach reflects a concerted intention to identify those resources that can now or in the future benefit from occupying an open data environment. Although the DCR as it now stands is not openly navigable, the addition of the PIDs together with the potential future facilitation of SKOS-enabled graph structures might well move the resource to a higher level on the spectrum. Nevertheless, a proposal has come forward to replicate at least a critical subset (presumably the core terminology subset) in the IEC's Common Data Dictionary (CDD) platform. Further options for standardizing or normalizing a select set of the data categories (e.g., those used in ISO standards in conjunction with ISO 10241) may eventually be realized in the framework of the current ISO SMART (ISO Standards Machine Applicable, Readable and Transferable) initiative or with reference to the W3C's *Data Catalog Vocabulary (DCAT)* recommendation (W3C 2020).

6.4 ‘Thing’-oriented collections

So-called ‘Thing’-oriented systems like the IEC’s CDD and the eCl@ss master data standard for products and services combine aspects of terminology management with taxonomical ordering to support electronic data exchange in the context of industrial supply chains. The CDD comprises IEC 61360-4, which provides a reference collection of standard data element types and component classes in the form of a well-structured classification system, accompanied by a detailed data set associated with the object specified in any given entry. However, the potential user is disappointed to find that instead of being able to output a serialized representation of these entries or to filter sets of entries, it is only possible to output individual entries one at a time in a convoluted, difficult-to-interpret set of Excel® files. Thus, the CDD is only a human-readable resource, and there is no apparent way to access this otherwise attractive resource in a machine-readable way, despite the fact that the companion standard IEC 61360-2 provides a formal model and a means for the computer-sensible representation and exchange of product-related data. In order to benefit from fully machine-sensible application of the data, the potential user must turn to eCl@ss, which provides a platform for product master data that can be exchanged digitally “across sectors, countries, languages and organizations” for over 45,000 product classes in support of product identification, procurement, and communication throughout a wide range of industries (eCl@ss). It is apparent that the primary effort to create and apply these standards has gone into this pay-to-license resource rather than allowing the serialization of the free-to-the-public CDD, with its relatively limited set of component classes.

It is not entirely clear why the thing-oriented CDD would be considered an attractive vessel for the content of the TC37 DCR, other than an effort to reclassify the current set in a standardized resource (where they would be locked into an inaccessible environment with respect to machine readability). It was at one point proposed to replicate the 12620-1 data model in the CDD, and not necessarily to replace those entries in the DCR, but rather to mirror them. Since the CDD is clearly built on the ISO 36160 data model, it was not clear if this proposal would be doable. Standardizing data category specifications would go through a meaningful standardization process, thus undergoing change, so it would become difficult to maintain the SSoT principle with such a dual arrangement. In the meantime, TC 37 has abandoned the option to standardize data categories in the CDD, but is considering standardization of those data categories used to create ISO standardized terminologies within the framework of the new ISO SMART system (ISO-SMART, 2022).

7. Outlook

The current version of ISO 704 is the first edition of the standard to explicitly spell out the position of the standard with respect to text-oriented terminology management, reflecting the conscious recognition that other approaches exist. Other studies, for instance Wright and Budin (2001), introduced topics such as information management, commercial and industrial applications, data retrieval, and corpus management, but these articles tend to stand alone without a strong sense of integrated interoperability, and, of course, the maturing vision of an open Web-based network of linkable information had yet to evolve. At this writing, directions in JTC 1/SC 32 project new standards involving terminological issues that extend beyond the scope of this chapter. More and more environments in industry involve the integration of more traditional discourse-oriented terminology management with thing-oriented systems associated with process engineering and logistics, informed on all sides by the opening up of systems in order to facilitate machine reasoning across Web boundaries. While there will always be a place for systems like eCl@ss that may remain ensconced behind their domain-specific paywalls, care should be taken both in standards and in system development to promote open systems with an eye to greater interactivity.

Abbreviations

The trend in this field is towards terminologization of abbreviated forms. In many cases full forms are rarely or never used.

AAAI	Association for the Advancement of Artificial Intelligence
ASTM	originally American Society for Testing and Materials, now ASTM Intl.
ATA	American Translators Association
CAS	originally Chemical Abstracts System
CCR	CLARIN Concept Registry
CDB	[ISO] Concept Data Base
CDD	[IEC] Concept Data Dictionary
CoP	Communities of Practice
DCAT	Data Catalog Vocabulary
DCR	Data Category Registry / Repository
DOL™	Distributed Ontology, Model, and Specification Language
IEC	International Electrotechnical Commission
IoT	Internet of Things
ISO	International Organization for Standardization
ISO/IEC JTC1	ISO/IEC Joint Technical Committee 1
LBX	Lexical Base eXchange

LISA/OSCAR	Localization Industry Standards Association/Open Standards for Container/ Content Allowing Re-use
LLOD	Linguistic Linked Open Data
LoC	(US) Library of Congress
LMF	Lexical Markup Framework
KWIC	Keyword in Context
MARC	Machine Readable Cataloguing
MRD	Machine Readable Dictionary
NLP	Natural Language Processing
OBP	ISO Online Browsing Platform
OASIS	Organization for the Advancement of Structured Information Systems
OMG	Object Management Group
OWL	Web Ontology Language
PID	Persistent Identifiers
RA	(ISO) Registration Authority
RDF	Resource Description Framework
RNG	RELAX NG, REgular LAnguage for XML Next Generation
SAE International	originally Society of Automotive Engineers
SC	Sub-Committee
SGML	Standard Generalized Markup Language
SIL International	originally Summer Institute of Linguistics
SKOS	Simple Knowledge Organization System
SMCC	[ISO] Smart Manufacturing Coordinating Committee
SPoT	Single Point of Truth
SSoT	Single Source of Truth
T9n	Translation
TBX	TermBase eXchange
TEI	Text Encoding Initiative
TC	Technical Committee
TEnTs	Translation Environment Tools
TM	Translation Memory
TMF	Terminological Markup Framework
TMS	Terminology Management System
UML	Unified Modeling Language
W3C	Worldwide Web Consortium
WG	Working Group
XLIFF	XML Localisation Interchange File Format
XML	eXtensible Markup Language

Concept management for Terminology

A Knowledge Engineering approach

Ingrid Meyer

University of Ottawa

Terminology has both a linguistic and a conceptual dimension, and the general goal of our research is to develop a computer aid to facilitate the latter. Concept management is crucial to all stages of terminology work as it is practiced today and will become even more so for the next generation of term banks. The central argument of this paper is that concept management technology should draw on developments in Knowledge Engineering, since knowledge engineers and terminologists share a number of fundamental conceptual problems. As an illustration, we describe the work underway at the Artificial Intelligence Laboratory of the University of Ottawa, where a knowledge engineering tool for terminology applications is under active development.

Keywords: Terminology, knowledge engineering, concept management, terminologist's workstation, term bank, terminotics, computer-assisted terminology

1. Introduction¹

Computers have been associated with Terminology² since the early 1970s, when they began to be used as repositories of terminological data. While storage continues

1. Reprinted, with permission from *Standardizing Terminology for Better Communication: Practice, Applied Theory, and Results*, ed. Richard Strehlow and Sue Ellen Wright, copyright ASTM International, 100 Barr Harbor Drive, West Conshohocken, PA 19428. Changes were made to the original version only to the extent that they complied with the formatting requirements of this volume.

2. The term *terminology* has traditionally been used in three senses: (1) the scientific study of terminology, (2) the practical activity of terminology work, and (3) the set of terms representing the system of concepts of a particular subject field. Following the recently published ISO International Standard 1087: Terminology-Vocabulary (1990), we shall use the term *terminology science* to designate Sense 1, *terminology work* to designate Sense 2, and *terminology* (e.g., a *terminology*,

to be of great interest, particularly in light of developments in CD-ROM technology, many researchers are now focusing on acquisition and processing.³ From this perspective, the research problem is to discover which aspects of the terminologist's day-to-day activities can be fully or partially automated and, subsequently, to design tools for a terminologist's workstation.

To date, almost all research and development for a workstation environment has focused on the linguistic dimension⁴ of terminology work – for example, on the linguistic aspects of tasks such as identifying terminological units (both simple and complex) in machine-readable texts, delimiting contexts, and establishing concordances. These important problems are attracting a good deal of research attention since they represent highly time consuming and repetitive aspects of terminology work, and also since the appropriate technology is, for the most part, already quite advanced. Ultimately, however, we feel that a terminologist's workstation must also provide support for the conceptual dimension of terminology work, which consists in acquiring, analyzing, comparing, formalizing, revising – in other words, managing – a variety of information about the concepts designated by the terms. Of course, conceptual problems are substantially more difficult than purely linguistic ones, and a suitable technology has not been available. Fortunately, recent developments in the subfield of Artificial Intelligence, known as Knowledge Engineering, are now making computer assistance for terminological concept management a feasible proposition.

This chapter has two sections. Section 2 provides an analysis of the conceptual dimension of terminology work in order to show that concept management is not merely an esoteric, intellectual activity, but rather one with important practical applications. This section also proposes that the conceptual dimension of terminology work will become even more crucial for the next generation of term banks. Section 3 argues that terminology-oriented concept management should draw on

terminologies) to designate Sense 3. Occasionally in this paper (e.g., the sentence to which this footnote applies) we wish to designate Senses 1 and 2 simultaneously; in this case, we have used the term *terminology*.

3. For a systematic overview of current research and development efforts in this area, see Auger, Drouin and L'Homme (1991).

4. We shall use the term *linguistic* (in opposition to *conceptual*) in the very general sense of "related to the terms, as opposed to the concepts designated by the terms." Of course, any attempt to separate terms from concepts, while useful for analysis purposes, is fundamentally artificial; terms and concepts are intimately linked. For this reason, we have used the term *dimension* (e.g., *linguistic dimension*, *conceptual dimension*), which we feel is less likely to evoke a sense of separation than other candidate terms such as *component*. This term is also used in Sager (1990). Sager's *cognitive dimension* corresponds to our *conceptual dimension*.

recent developments in Knowledge Engineering since knowledge engineers and terminologists share a number of fundamental conceptual problems. To illustrate our point, we describe work underway at the Artificial Intelligence Laboratory of the University of Ottawa, Canada, where a knowledge engineering tool for terminology applications is under active development, based on initial testing in two terminology environments.

2. The role of concept management in terminology work

2.1 Concept management/or today's terminology work

Conceptual activities permeate virtually all aspects of Terminology. Since the nature and relative importance of conceptual work depends on the type of terminology work,⁵ the following discussion is restricted to thematic,⁶ bilingual terminology work as it has traditionally been practiced in the Canadian context (overviews of the Canadian approach to Terminology can be found in Dubuc (1985) and Rondeau (1984); an overview of the Secretary of State's approach can be found in Cosette (1976)). Our discussion is also influenced by the fact that most terminologists in Canada are not normally true subject-field experts (their training is typically in Translation, Linguistics, etc.), although they can be extremely knowledgeable in their fields.

In the following section, we analyze four interrelated⁷ categories of terminological activities: selection of documentation, establishment of a terminology, preparation of term records, and quality control. We briefly explain each general activity, describe its conceptual subactivities, and then point out the practical applications of these conceptual subactivities. A summary of our analysis is provided in Table 1.

5. For an overview of the conceptual aspects of a variety of terminology activities, see Meyer, Miller and Michaud (1991).

6. By *thematic terminology*, we designate the activity of collecting and describing (as exhaustively as practical constraints allow) the terminology of a given field or subfield. This is in contrast to terminology work on isolated terms (this kind of terminology work is done at the Secretary of State in response to telephone requests, for example). A distinction can be made between *open* and *closed* thematic terminology; the first refers to terminology work which is not based on a preestablished terminology, while the second refers to terminology work that is based on a term list supplied by a client. This paper is most directly related to thematic terminology work of the open type.

7. None of these categories of activities can be practiced in isolation from the other activities: for example, the selection of documentation can be subsequently modified; quality control can (and often, should!) be carried out at various stages of a project, etc.

Table 1. The conceptual dimension of terminology work

General activity	Conceptual activities	Practical applications
<i>Selection of documentation</i>	Initial familiarization with knowledge structures	Clarification of project scope
	Initial familiarization with characteristics of key concepts	Orientation of search for documentation
		Judgment of quality of documentation
		Communication with documentation resource persons and experts
<i>Establishment of a terminology</i>	Analysis of knowledge structures	Identification of simple and complex terms
	Analysis of conceptual characteristics	Identification of ‘noise’ and ‘silences’
		Evaluation of quality of contexts
		Identification of obvious synonymy
<i>Preparation of term records</i>	Continued analysis of knowledge structures	Continued communication with documentation resource persons and experts
	Continued analysis of conceptual characteristics	Attribution of subject-field labels
	Comparison of conceptual characteristics	Definition construction
		Context selection
		Identification of synonymy
<i>Quality control</i>	Analysis of specific conceptual problems	Establishment of interlinguistic equivalence
	Monitoring of changes in knowledge structures and conceptual characteristics	Creation of neologisms
		Communication with experts and other terminologists
		Comparison between new and previous conceptual characteristics and knowledge structures

2.2 Selection of documentation

General activity – Before any collection or analysis of terms can occur, terminologists must select the documentation corpus that will serve as the principal ‘knowledge source’ for the project. As far as possible, terminologists will also elicit the collaboration of subject-field experts, and possibly other terminologists, as additional knowledge sources.

Conceptual activities – When they are assigned to a new project, terminologists typically begin by acquiring some general knowledge about the subject-field by doing introductory reading in textbooks, encyclopedia articles, popular journals, and so on; in effect, they are laying the conceptual foundations for the project. Conceptually speaking, this initial familiarization process with the field occurs at both a macro- and a microlevel.⁸ On the macrolevel, terminologists begin to

8. The distinction between macro- and microlevels of conceptual analysis, while useful to our discussion, must nevertheless be seen as fundamentally artificial. The macrolevel (which we

familiarize themselves with the general knowledge structures of the field, trying to determine its boundaries, subdivisions, and areas of overlap with other fields (for multidisciplinary fields). Often, at this stage, terminologists will sketch out these 'skeletal' knowledge structures in the form of a concept network. On a microlevel, terminologists will pay special attention to a number of individual concepts which, in the course of their reading, emerge as being particularly important to the field.

Practical applications – These preliminary conceptual activities are crucial to documentation selection since they help to clarify the project's scope; a clear idea of the conceptual boundaries of the field helps delimit the range of documentation to be sought. Determining areas of overlap with other fields also helps terminologists establish links with related documentation. Once the general scope is established, the extensive documentation search can begin. Here, a clear idea of the major subfields helps terminologists orient their work along a number of documentary 'paths.' These paths may be prioritized, for example, in cases where users have an immediate need for terminology in a particular subfield. Familiarization with the field's macro- and microstructures helps provide specific points of entry into the documentation since the names of subfields, of key concepts, and of the characteristics of these concepts can be used as keywords for the search. Having a general idea of the hierarchical structure of the field also helps orient the process of documentation selection since terminologists tend to proceed from general to more specific literature.

Once a certain mass of potential documentation is obtained, their general knowledge of the subject field provides terminologists with a yardstick for judging the quality of this documentation. It also helps terminologists classify their documentation according to subfield. In the Canadian context, this is particularly important since translators work bilingually; to 'manage' the large amounts of documentation to be scanned (see next subsection), they very often scan one subfield at a time, in one language and then in the other, before proceeding to another subfield.

Finally, these preliminary conceptual activities provide terminologists with the conceptual framework and basic terminology needed for effectively communicating with librarians and other documentation resource persons, as well as with subject-field experts. Communication is particularly important in the case of experts, who tend to be very busy; having done their 'homework,' terminologists will be able to direct the conversation in order to elicit maximum information in a minimum amount of time. In a nutshell, "starting out on the right foot" in this way boosts terminologists' credibility with experts and increases their chances of convincing experts to remain involved as the project advances.

associate with the expression *general knowledge structures*) is shaped by the conceptual characteristics of individual concepts (i.e., the microlevel), since these characteristics indicate relations to other concepts.

2.3 Establishment of a terminology⁹

General activity – Once the documentation has been selected, it undergoes a process called scanning, i.e., careful reading, with the extraction of potential terms along with their contexts. Additional research may be needed for specific problems (e.g., terms not found for concepts identified, terms with inadequate contexts), after which the data is organized by grouping the various instances of a term, noting obvious cases of synonymy, abbreviations, usage labels, etc.

Conceptual activities – Through the scanning process, terminologists begin to analyze¹⁰ the general knowledge structures of the field, fleshing out (whether on paper or in the mind) the skeletal concept network drafted during their background reading. They also begin to analyze the conceptual characteristics of individual terms (i.e., the terms found in the documentation), based on the contexts in which the terms appear.

Practical applications – Based on their general understanding of the subject field, terminologists will be able to begin identifying the lexical items that constitute terms within the field. An understanding of conceptual relations will help in identifying complex terms. The process of term identification involves eliminating terms that would constitute ‘noise’ in the terminology, i.e., lexical items that are simply non-terminological or terms that do not fall within the established boundaries of the field. As well, terminologists will want to identify lacunae in the preliminary terminology. Conceptually speaking, this implies that no terms have yet been found for certain areas of the subject-field. When the time comes to organize the data (terms and contexts), their understanding of conceptual characteristics will enable terminologists to judge the quality of contexts, eliminate those that are not adequate, and seek additional contexts where necessary.

As is pointed out in *Le dépouillement en terminologie* (1980), terminologists tend to err on the side of caution with regard to eliminating contexts, since their understanding of the pertinent concepts is still not very deep. They will also be able to identify some cases of synonymy, which will help in the regrouping of terms. Finally, as terminologists prepare to finalize the terminology (i.e., to determine the terms for which records will be prepared), the conceptual framework acquired so far will help them continue to communicate about problem areas with documentation resource persons and subject-field experts.

9. The term *base list* is used at the Secretary of State to describe this preliminary terminology.

10. We say “begin to analyze” because analysis is still not very deep at this stage, given that scanning involves rather fast reading. The cognitive processes that occur at this stage are not yet very well understood, but we are assuming that some level of analysis, even if it is partly subconscious, does occur at this stage.

2.4 Preparation of term records

General activity – Using the established terminology and associated contexts, terminologists can begin a systematic analysis of terms-in-context. The primary function of this analysis is to determine the meanings of the terms, although it also serves to identify other linguistic characteristics such as part of speech, gender, frequency, geographic origin, etc.

Conceptual activities – The terminologist’s conceptual goal at this stage is to achieve the depth of understanding needed to complete the term records. On a macrolevel, this implies establishing the main conceptual partitions of the field and also determining hierarchical and other relations as far as possible. On a microlevel, contextual analysis will involve identifying a certain number of conceptual characteristics for all concepts. These characteristics will then be compared with those of potentially related concepts (e.g., synonyms, target-language equivalents) in order to determine those which are necessary for establishing a conceptual match.

Practical applications – A proper understanding of the conceptual structures of the field is crucial to determine the subject-field label for a term. It is also crucial to determine cases where a single term is used in more than one subject field to designate the same concept (in order to prevent a proliferation of records); when several subject fields apply, terminologists will want to indicate the most ‘basic’ one, which again implies a certain depth of understanding of the conceptual structures of the fields in question.

Probably the most important application of conceptual analysis is definition construction. Whether a definition is hand-crafted or selected from a context, terminologists will need to understand the concept’s characteristics. If they are attempting the classic intensional (i.e., genus-differentia) definition, terminologists will then need to compare these conceptual characteristics with those of concepts at the same hierarchical level (i.e., with the characteristics of the coordinate concepts)¹¹ in order to determine the distinguishing characteristics (i.e., the differentia in an intensional definition). Relations other than the generic-specific (e.g., whole-part, cause-effect, tool-function) may also be analyzed and reflected in definitions. When a contextual definition is used, a sound understanding of conceptual characteristics and relations will help the terminologist select the most useful context (i.e., the one that provides the greatest number of key characteristics).

Identifying synonyms requires a careful comparison of conceptual characteristics in order to determine that these are indeed identical for the terms in question.

11. According to the ISO International Standard 1087: Terminology-Vocabulary, a coordinate concept is a “concept in a hierarchical system which ranks at the same level as one or more other concepts”.

When two concepts differ in only a very few (and not very significant) characteristics, they may be designated as pseudosynonyms (e.g., one concept may have one more characteristic than another and thus be more specific). Establishing a conceptual match is also crucial to multilingual terminology work, which is complicated by the fact that conceptual structures often do not correspond perfectly from one language to another, resulting in cases of incomplete equivalence. Sometimes there may be no equivalent in the other language at all, resulting in the need to create a neologism. While we cannot fully describe the complexities of this problem here, suffice it to say that conceptual analysis is essential for determining whether the concept already exists within the current knowledge structures of the target language, and when it does, what its characteristics are (since the concept is so new, its characteristics, and consequently its location within the knowledge structures, may still be fluctuating). In many cases, an existing term will be adopted to designate the new concept, and conceptual analysis of the candidate terms is essential for determining which one possesses the greatest semantic compatibility with the new concept.

2.5 Quality control

General activity – Quality control can be achieved by two types of activity: revision and updating. On the one hand, before the project is completed, the various types of information collected by the terminologist should be revised by subject-field experts or by other terminologists (e.g., terminologists with experience in neighboring or related fields or more experienced terminologists). On the other hand, after the project is completed, a periodic updating of terminology records can occur whenever this is justified by changes and expansion in the subject field.

Conceptual activities – Revising the results of a terminology project involves analyzing and discussing specific conceptual problems identified by the experts and/or other terminologists. Periodic updating implies a monitoring of changes in knowledge structures and conceptual characteristics.

Practical applications – To facilitate revision, terminologists need a sound understanding of the subject field in order to interpret feedback from experts and to elicit information on this feedback (e.g., when terminologists do not understand feedback, when the feedback contradicts what the terminologists found, or when experts give conflicting feedback). Regarding updating, a clear understanding of the current state of the knowledge will give the terminologist a basis for comparison when new structures and conceptual characteristics emerge. Of course, conceptual problems increase when a field is particularly large or has complex knowledge structures, or when the field is changing rapidly.

2.6 Concept management for the next generation of term banks

As illustrated above, conceptual problems pervade all aspects of terminology work as it is currently practiced; however, we feel that concept management technology will become even more crucial in the next generation of term banks. The following are just some of the factors that will contribute to this trend.

Number of concepts – There is no sign that the current information explosion is a plateau; on the contrary, the number of concepts will probably continue to rise rapidly. New concepts are typically related to previous concepts and knowledge structures; terminology banks that are describing conceptual ‘skyscrapers’ will want to make sure that their conceptual ‘foundations’ are solid.

Multidisciplinarity and complexity – Multidisciplinarity and complexity continue to characterize modern science and technology, increasing the terminologist’s need for assistance in maintaining conceptual clarity and consistency.

Demand for quality – An increasing demand for quality continues to characterize the fields served by term banks, particularly the various activities that comprise the document-production chain; in an increasingly competitive market, companies are realizing that high-quality documentation can be a factor in a product’s success.

Number of users – There are many reasons why the number of users is anticipated to increase significantly. On the most practical level, to justify high costs, term banks are seeking to attract more users through remote log-in and the availability of data on CD-ROM. The rapid pace at which new concepts are developing makes these options increasingly appealing: many language professionals are finding it impossible to reconcile personal terminology work with heightened pressures to speed up output. The increasing multidisciplinarity of subject fields is having a similar effect. Specialized translators, for example, are more and more often confronted with texts containing terms from a variety of fields and do not have the time (or the expertise) to do the appropriate terminology work themselves.

Types of users – As user numbers increase, so will user types. While, in the past, translators have been the principal user category, it is anticipated that increasing numbers of users will come from the ranks of subject specialists, special-subject learners, technical writers, documentation specialists, language planners, and so on. Since these users will differ in their subject-field expertise, terminology banks will require more knowledge and highly structured knowledge so that the knowledge can be properly graded according to user profile, and so that a large number of entry points into the knowledge can be possible (see data retrieval, below). It is predicted (e.g., Sager 1990) that machines may become a category of user for terminology banks; machine translation tools, expert systems, natural-language interfaces to databases, and spelling checkers are just a few of the most obvious applications

that have been suggested. Machine use will dramatically increase the importance of concept management, placing great demands on precision and formalization. Machines will need very large quantities of explicitly represented conceptual information since they do not possess much of the basic real-world information that humans know implicitly.

Data retrieval – The large diversity of users of terminological data will inevitably be accompanied by demands for more varied mechanisms of retrieving this data. Developments in hypertext are already encouraging this trend in information retrieval systems in general. While traditional term banks have been largely limited to linguistic entry points to the data, the next generation will need to supply conceptual entry points as well. For example, users will want to be able to ask questions like “what do you call the machine that accomplishes X?”, “what do you call the material that has physical properties X, Y, and Z?”, and so on. In other words, many of the conceptual relations that are implicit (often in fragments) in definitions and contexts will need to be made explicit and more complete.

Multilingualism – The increasing transfer of information across national and linguistic boundaries is creating a heightened need for high-quality bilingual and multilingual term banks that will facilitate the production of documentation in various languages. The more languages that are added to a term bank, the greater the conceptual complexity of the task. Sager (1990) has predicted that computers are making terminology work a “discontinuous process”; for multilingual work, this implies the distribution of terminology work over countries and continents so that, for example, native speakers can carry out the relevant work in their own linguistic environment. The resulting ‘discontinuity’ will need to be counter-balanced by rich and formally structured knowledge sources, implying sophisticated concept management environments.

3. A Knowledge Engineering approach to concept management

Despite the practical applications of conceptual activities for term banks of the present and future, terminologists still do not enjoy the support of computer assistance for the conceptual dimension of their work. Typically, they employ a combination of ‘paper-and-pencil’ and ‘do-it-in-my-head’ methods of concept management, which, as discussed by Skuce and Meyer (1990), entail problems of storage and retrieval, validation and updating of information, shareability of notation, and maintenance of consistency and clarity. While computer assistance is clearly desirable for the conceptual dimension of terminology work, it has only recently become feasible with developments in an emerging subfield of Artificial Intelligence called *Knowledge Engineering*. In the following sections, we analyze the fundamental

conceptual similarities between the work of the knowledge engineer and that of the terminologist (see next section), and subsequently describe a knowledge engineering tool that we are currently developing with terminological applications in mind (see CODE: A Knowledge Engineering Tool for Concept Management).

3.1 Similarities between Knowledge Engineering and Terminology

Knowledge Engineering is grounded in the critical need for technology and methodology to build systems that are knowledge-based (as opposed to algorithm-based). As the name implies, knowledge-based systems need to ‘know’ (or at least ‘represent’) what human experts know, since they are designed to solve problems in specialized fields of human expertise. Expert knowledge in these systems has two components: a corpus of factual subject-field knowledge and a set of strategic rules for applying this knowledge to problem solving. While both components are important to Knowledge Engineering in general, we shall limit our discussion to questions of factual subject-field knowledge, which is particularly relevant to Terminology.

At the heart of the relationship between Knowledge Engineering and Terminology is the fact that practitioners of both disciplines function as intermediaries in a knowledge communication context involving experts on the one hand and a given technology on the other. There are practical differences, of course – knowledge engineers typically acquire most of their knowledge from experts directly, while terminologists acquire most of it indirectly through documents written by experts (though, as mentioned above, the active involvement of experts is an essential component of a terminology project); knowledge engineers store the acquired knowledge in a knowledge base (i.e., a repository containing only knowledge), while terminologists store it in a terminology bank that contains conceptual as well as linguistic information. Despite these and other differences, however, knowledge engineers and terminologists share a number of fundamental conceptual problems. In the following sections, we describe three important subtasks of Knowledge Engineering: knowledge acquisition, knowledge formalization, and knowledge refinement, and, for each, analyze the points of similarity with Terminology.

Knowledge Acquisition – Knowledge engineers can elicit knowledge from experts or extract it from documents, although elicitation is usually preferred. The problems of eliciting expertise from a skilled person are well known in psychology (Gaines 1990). One such problem is that experts often have trouble expressing their knowledge clearly. According to Gaines (1990), expertise consists of three elements – performance, understanding, and communication – that require the expert to play the roles of practitioner, scientist, and teacher, respectively. Unfortunately, experts vary widely in their teaching skills. Intimately linked with teaching skills are language skills; experts may not be able to transmit expertise efficiently because

they are unable to express it clearly. Another problem of expertise elicitation is that experts may not provide exactly the knowledge that is required: some elements of information may be irrelevant, while others may be neglected. All of these problems are aggravated by the fact that knowledge engineers, like terminologists, are often not subject-field experts, but rather, are typically trained in Computer Science.

Knowledge acquisition, which has come to be called the *bottleneck* of Knowledge Engineering, is an even greater problem in Terminology in the sense that terminologists rely heavily on documents for their knowledge; when these show evidence of the knowledge communication problems discussed above, terminologists are not usually able to discuss the problems with the authors of the texts. Of course, when they do communicate directly with experts, terminologists face the same problems of unclear, irrelevant, and incomplete knowledge that confront the knowledge engineer.

Knowledge formalization – As knowledge engineers acquire knowledge, they face the challenge of formalizing (i.e., ‘engineering’) it so that it can be applied to a given purpose. Formalized knowledge stored in a computer is usually called a *knowledge base*. Complicating knowledge base construction is the fact that the knowledge must be usable by humans (e.g., clients, subject-field experts, knowledge engineers) and machines (e.g., the knowledge-based system and systems that interact with it). Unfortunately, knowledge “does not come off the shelf, prepackaged, ready for use,” in the words of Hayes-Roth (1984). Rather, human knowledge is characterized by extreme diversity. For example, expert knowledge can be inconsistent and contradictory, as the well-known term *expert opinion* suggests. It can also be multidimensional, since people’s understanding of a concept or conceptual system can depend on their particular point of view. And finally, knowledge can be hard to ‘capture,’ since it is constantly changing, and since emergent knowledge can be incomplete and unclear.

Inconsistency, contradiction, multidimensionality, change, incompleteness, and lack of clarity also plague terminologists as they try to formalize expert knowledge for the various practical applications outlined in the previous section, entitled “Concept management for today’s terminology work.” These problems will become even more serious impediments to knowledge management for the next generation of term banks; if the vision of a variety of both human and machine uses becomes a reality as described in the previous section, then term banks will, in effect, increasingly come to resemble knowledge bases.

Knowledge refinement – Knowledge engineers can refine knowledge, once formalized, in two ways. On the one hand, they may validate it by testing the knowledge-based system on the intended application and subsequently revise aspects of the knowledge. On the other hand, they may periodically update the

knowledge, for example when the knowledge engineer's understanding of the field deepens or expands as a result of more knowledge acquisition or formalization activities, when the field itself changes, or when the system needs more knowledge due to changes in the application.

For the terminologist, knowledge refinement does not involve system testing (although this may occur in the future if term banks begin to serve machine users). However, as described in the previous section, entitled "Quality control," it requires revision by subject-field experts and other terminologists, as well as periodic updating as knowledge changes and as the terminology is used by clients. These quality control issues will become more important as more quality is demanded by users.

Given the large amounts of knowledge required by 'intelligent' systems, developing tools to assist in acquiring, formalizing, and refining this knowledge has become a research priority in Knowledge Engineering. One such tool, called CODE (Conceptually Oriented Descriptive Environment), has been under active development for about three years at the Artificial Intelligence Laboratory of the University of Ottawa, Canada. During the first phase of the project, which has just ended, a working prototype was developed and tested in two terminologically oriented pilot research projects, one with the Terminology and Linguistic Services Directorate of the Secretary of State of Canada and another with Bell Northern Research (the Canadian equivalent of Bell Labs).

This first phase of our research has already been well documented elsewhere; a general technical description of the CODE system can be found in Skuce, Wand and Beauvillé (1989) and Skuce (1989); a description of those features that have proven particularly useful for Terminology can be found in Skuce and Meyer (1990); Meyer and Paradis (1991) and Meyer, Miller and Michaud (1991); descriptions of the CODE project at the Secretary of State and Bell Northern Research can be found in Skuce and Meyer (1990) and Meyer and Paradis (1991); and finally, methodological considerations are discussed in Skuce and Monarch (1990).

Our research is now entering a second three-year phase, which has three goals: first, to develop a new version of the CODE system that will incorporate ideas stemming from the two practical experiences at the Secretary of State and Bell Northern Research; second, to develop a methodology to support the new version of the system; and third, to use the technology and methodology to construct a prototype terminology bank (which we call a *terminological knowledge base*) with a rich and highly structured conceptual component. Below, we give an overview of the technological framework for our current research activity. Within this framework, the following features are receiving particular attention.

User interface – We are aiming at an interface that is flexible enough for a wide range of contexts. For example, it should suit terminologists as well as other

language professionals who are required to describe terms in the course of their work (e.g., translators, technical writers, product designers).¹² It should also be useful to persons with various levels of subject-field expertise, from true experts, to terminologists who have been working in a field for a long time, to terminologists starting out in a subject-field. It should be useful as a communication tool, for example as a basis for discussions between terminologists and experts, between terminologists working as a team on a project, between members of standardizing committees, etc. And finally, it should be useful as a teaching tool (e.g., for new terminologists, students of special subjects, translation students).

Graphics – The graphical display of the current CODE systems is already very well developed, and we intend to enhance it even more in the future since terminologists place an extremely high value on graphical representations of knowledge structures. For example, even using the paper-and-pencil method, terminologists often find it helpful to draw concept networks of their fields or subfields. We are aiming at sophisticated mechanisms for (a) focusing on certain parts of graphs, (b) comparing and contrasting multiple graphs, (c) isolating ‘special’ (e.g., fuzzy, uncertain, unconfirmed) concepts or groups of concepts, (d) representing multidimensionality and fluidity (see below), and (e) showing both hierarchical and nonhierarchical relations. We would also like to be able to incorporate images of all kinds (e.g., not just graphs).

Hypertext browsing – This feature is already present in the current version of the system and will be developed even further. When a subject field is very large, multidisciplinary, or rapidly changing, when multiple knowledge bases are consulted simultaneously (see below), and when the knowledge base is used as a communication or learning tool, quick and easy navigation through the data is crucial. We intend to support browsing through a global lexicon facility that could search on any word in the knowledge base, and through a capacity for fuzzy (i.e., approximate) searching.

Multiple knowledge bases – Terminologists would want to consult multiple knowledge bases simultaneously when they are working in interdisciplinary or overlapping fields and also when they are doing multilingual work (since knowledge structures rarely correspond from one language to another). We are developing support for isolating areas of correspondence, non-correspondence, and

12. This type of ‘Terminology’ is what we have described elsewhere (Meyer and Skuce 1990) as terminology as part of the document production chain, i.e., terminology seen in its broadest sense as just one component of a wide spectrum of document production and dissemination activities including, for example, product design specification, technical writing, revision, proofreading, translation, abstracting, management information, and marketing. Conceptual and linguistic consistency is crucial throughout all phases of this chain of activities.

overlap, and also for generating parts of knowledge structures automatically (e.g., in multilingual work, when certain parts of one knowledge base do correspond quite well between languages). Ultimately, this will involve a machine translation component.

Multidimensionality – Terminologists are all too aware that fields can frequently be partitioned in many different ways, depending on which conceptual characteristics are stressed. We intend to support the representation of multidimensionality through a variety of inheritance¹³ features as well as through graphics.

Fluidity – In the words of Sager (1990), “knowledge structures are not absolute ... conceptual systems are relatively fluid entities constantly undergoing change.” We are working on mechanisms for allowing easy modification of knowledge structures as they change, or as the terminologist’s understanding of these structures deepens. Changes to conceptual information will be monitored through mechanisms for quality control, discussed below.

Quality control – Quality control is important when fields are large, multidisciplinary, and rapidly evolving, and when conceptual information is being entered in a multiuser environment. In these situations, there is a need for consistency checking, which in the CODE system is intimately linked with inheritance mechanisms; the system queries the user whenever conceptual characteristics are modified or when concepts are added or moved. We are also developing a facility for doing ‘what-if’ experiments when the terminologist is unsure about a concept’s position within a knowledge structure.

Formality continuum – Some concepts and conceptual systems lend themselves very well to being ‘captured’ in a formal way. Scientific nomenclatures, for example, are traditionally structured in strict hierarchies. Fields outside of science and technology, however, can be much more difficult to structure formally, as Cole (1991) has pointed out. Formal representations are also difficult for newly emerging fields and subfields, where concepts are typically rather ‘fuzzy.’ The concept management technology we are aiming at should allow the terminologist to work in a ‘casual’ as well as in a ‘formal’ mode, to switch easily between modes depending on the application, and to ‘blend’ modes.

Support of definition construction – Since a definition is, in the words of Martin (1988), a “knowledge representation,” we are aiming at some level of automatic

13. Inheritance is a powerful AI technique that applies when concepts are arranged in a generic-specific hierarchy; inheritance allows any characteristic of a given concept to be implicitly true for all specializations (i.e., subordinate concepts) of this concept, for all specializations of these specializations, and so on. Thus, when a terminologist knows that a given concept belongs to a known class, he can automatically access a set of additional (i.e., inherited) knowledge about this concept.

definition construction based on an automatic comparison of characteristics of coordinate concepts (i.e., concepts at the same hierarchical level) to determine the distinguishing characteristics.

Acknowledgements

We are grateful to Line Paradis and Douglas Skuce for their critical reading and commenting of this paper.

Funding

The development of the CODE systems for terminological applications is being supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) and by Research Services of the University of Ottawa. Testing of the CODE system in terminology environments was supported by the Terminology and Linguistic Services Directorate of the Secretary of State of Canada and by Bell Northern Research. Douglas Skuce's general development of the CODE system is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the University Research Incentives Fund (URIF) of the Government of Ontario, Bell Northern Research, and Research Services of the University of Ottawa.

Multidimensionality

Lynne Bowker
University of Ottawa

Multidimensionality is a phenomenon of conceptual classification that arises when concepts are classified in more than one way within a concept system according to different characteristics. Because multidimensionality can be complex and cause information overload, early terminologists focused on the most common dimension. In the 1990s, knowledge-based approaches emerged, partly inspired by cognitive science models. Technological advances offered the possibility of managing multiple inheritance and generating graphical representations, making it easier to work with multidimensionality. Corpus-based techniques that used lexical knowledge patterns to uncover conceptual relations soon followed. Terminologists have thus incorporated multidimensionality into successive terminological knowledge bases, offering richer representations of multiple dimensions and a more holistic understanding of concepts. These resources also explore ways of addressing information overload.

Keywords: classification, characteristics, conceptual relations, concept systems, dimensions, graphical representation, Knowledge-based Terminology, multiple inheritance, terminological knowledge bases

1. Introduction

Imagine that you have a white T-shirt. You probably keep that T-shirt in a drawer along with your red T-shirt and your black T-shirt. You keep all the T-shirts together because, well, they belong together. Colour is not important here; they are all T-shirts, and they all get folded in a similar way and then stored together in your T-shirt drawer. But what happens on laundry day? Does your white T-shirt go into the washing machine at the same time as your red T-shirt? Or do you need to sort the T-shirts into different piles in order to wash them at different temperatures? Now colour *is* an important consideration. In this context, you no longer keep all your T-shirts together; on laundry day, the white T-shirt goes in the pile of white clothes to be washed in hot water, while the red T-shirt goes into the pile of coloured

clothes to be washed at a lower temperature. This, in a nutshell, is multidimensionality. The very same item – a white T-shirt – is sometimes viewed as a T-shirt and other times viewed as a white article of clothing. Maybe you have other ways of viewing it too. Maybe it is also a comfortable shirt (because it is made of soft fabric), or maybe it is an undershirt only worn beneath a dress shirt. Most items can be viewed from multiple perspectives. They have many qualities or characteristics, and at a given moment, one of these characteristics may be more important than the others. On laundry day, the colour is the most important characteristic. Once the laundry is clean and ready to be put away, another characteristic, such as “foldability” might take on greater importance.

These general ideas apply in Terminology too. Terminologists usually begin their work by identifying concepts in a specialized subject field. To understand a concept, the terminologist needs to identify its characteristics. The sum total of these characteristics makes up the concept’s intension. These characteristics also allow the terminologist to understand how the concept is related to other concepts (e.g., concepts can be grouped together because they share certain characteristics, or concepts can be separated from one another because they do not share certain characteristics). Together, information about a concept’s characteristics and relations can be used to create a definition for the concept, and this definition forms a bridge between the concept and the term. If a terminologist is working in more than one language, that definition also helps to establish interlingual equivalence: if the term in language A and the term in language B have the same definition (i.e., they refer to the same concept), then they can be considered equivalents and used in a task such as translation.

So where does multidimensionality come in? As explained by Kageura (1997b, 120):

It is generally accepted in terminological studies that a concept, or a unit of thought, consists of a set of characteristics (Felber 1984; Sager 1990). [...] Since the characteristics of a concept are frequently specified from different points of view or facets (function, material, shape, weight, etc.), a set of characteristics that constitutes a concept is normally multidimensional. From this point alone, we can expect a concept system to be multidimensional.

Yet, if multidimensionality is so fundamental to Terminology, why do we not often see it examined in detail? What challenges does it present for terminologists? In this chapter, we will explore the notion of multidimensionality in greater depth, but first it is worth providing some context about why this phenomenon has not always been a major feature of Terminology Theory and practice (Section 2). After exploring Terminology’s early years and learning why multidimensionality was largely absent during this period, in Section 3 we will then situate the beginning

of multidimensional research and practice in the context of Knowledge-based Terminology, which emerged in the 1990s. As part of this discussion, we will introduce some ideas from Cognitive Science (e.g., prototype theory) and Computer Science (e.g., Knowledge Engineering), which helped to set the scene for knowledge-based approaches to Terminology, including multidimensionality. In particular, we will present a number of guidelines that emerged as part of these initial forays into investigating multidimensionality and representing it in terminological knowledge bases (TKBs). Additionally, in Section 4 we will briefly consider how corpus-based approaches to Terminology have contributed to understanding more about multidimensionality. Next, in Section 5 we will present three different projects which have integrated multidimensionality into the construction of terminological resources (e.g., by representing multidimensional concept systems, by incorporating flexible definitions based on different dimensions or by using corpus-based techniques such as knowledge patterns to identify different dimensions). By following the progress of these various projects, we can see how they built on one another to identify and solve different challenges presented by multidimensionality. Finally, in Section 6 we will end with some concluding remarks and reflect on future directions.

2. Early years in Terminology: Physical objects, textbooks, linguistic orientation, and technological limitations

From the vantage point of the twenty-first century, it could be easy to look back and wonder why it is taking terminologists so long to come to grips with multidimensionality. In truth, however, this notion is a lot trickier than simply sorting your laundry, and even with all the knowledge gained over the past century, along with the support offered by modern-day technology, multidimensionality still presents challenges.

When Eugen Wüster first began to formalize his ideas for the General Theory of Terminology (GTT), he identified terminology as being an essential element of specialized communication (Wüster 1970 [1931]). To ensure that specialized communication was clear and precise, Wüster proposed an onomasiological or concept-based approach, and he also advocated for the principle of univocity, meaning that each concept should be designated by one term only, and each term should refer to only one concept (see chapters by Humbley and Candel in this volume). It is important to note that Wüster first developed his ideas while considering subject fields in which physical objects prevailed; this was the case for his early work on the machine-tool dictionary (Wüster 1968). The GTT places an emphasis on establishing a system of concepts where each individual concept in a subject

field can be fixed in a single and “correct” position within the concept system. When it comes to organizing concept systems, the concepts in fields like machine tools lend themselves to being organized into hierarchies, such as generic-specific hierarchies (also known as hyponymy or *is-a* or *type-of* hierarchies) or part-whole hierarchies (also known as meronymy or *has-a* or *part-of* hierarchies) (see chapter by Nuopponen in this volume). In a hierarchical concept system, the superordinate concept is divided into subordinate concepts according to a selected characteristic – usually one that is considered essential to the definition of that concept.

For example, we could divide the superordinate concept WINE into the subordinate concepts RED WINE and WHITE WINE according to the characteristic “colour”. If we divide the concept WINE according to “colour”, we could say that this represents one dimension of the concept. In the case of WINE, “colour” seems to be a very obvious choice for subdividing this concept. Indeed, in many subject fields a similarly obvious characteristic emerges as one that can be handily used to subdivide the superordinate concept. However, this does not mean that it is the only way of categorizing those concepts, or that it is the best choice for every situation. For example, another way of dividing WINE could be according to the characteristic “country of origin”, producing subordinate concepts such as FRENCH WINE and ITALIAN WINE. This represents a different dimension of the concept WINE and it represents a valid alternative classification. Yet another way to classify WINE is according to the characteristic “sugar content”, which produces a third dimension containing the subordinate concepts DRY WINE and SWEET WINE. Nonetheless, one of the characteristics often does emerge as being the most obvious or sensible choice in the majority of situations. We might even say that it is a prototypical choice, and we will return to this idea in more detail in an upcoming section.

In the meantime, since Terminology was such a new discipline in the mid-twentieth century, there was a lot of foundational groundwork to be laid. Commenting on the state of Terminology in Canada near the end of the 1960s, Delisle (2008, 124) observes that Terminology was not taught; textbooks did not exist; terminological research methods were embryonic; computerized term banks had not been developed; and the tasks of terminologists had yet to be defined. Nonetheless, driven in part by the needs of the rapidly expanding translation industry, terminology training was clearly needed, and so it began to be offered as part of translation education programs in Canada (and elsewhere) starting in the 1970s (Delisle 2008, 273). Of course, where there is teaching, there is also a need for pedagogical materials. Therefore, it was also during the 1970s and early 1980s that the first serious efforts to develop and document a methodology for conducting terminology work took place, leading to seminal works such as those produced by Auger and Rousseau (1978); Dubuc (1978); Wüster (1979); Rondeau (1984); Felber (1984), and Picht and Draskau (1985). These pioneering authors had the

difficult task of working out the basic principles and methods of the new discipline of Terminology, and of presenting them in a way that was easy for others to learn. Therefore, the adage “learn to walk before you run” certainly applies here. Multidimensional concept systems are complex and messy, and so it would not make sense to begin by tackling multidimensionality first. Rather, as noted above, there is often an obvious or prototypical classification that emerges, or a classification that may not be obvious but on which scientists agree; therefore, it was logical to begin by developing initial methods and pedagogical materials that took such relatively clear-cut or agreed-upon classifications as a basis.

Another important consideration to keep in mind is that, for Wüster, one important function of Terminology was to help *standardize* the language used in specialized fields of knowledge in order to ensure clarity and precision in communication. Therefore, while the starting point was the concept, the overriding goal of Terminology in the early days was more linguistic in nature. Indeed, this general trend continued for many years and is still reflected in the design and content of many well-known term banks that exist today, such as the Government of Canada’s TERMIUM Plus or the European Union’s Inter-Active Terminology for Europe (IATE). The principal users of these term banks are translators whose job is to produce texts using specialized language. As a result, much of the information contained in the term banks is linguistic, although we can find some information about conceptual relations represented in a more implicit way (e.g., through the indication of the broader subject field to which a concept belongs, or as part of an intensional definition that makes reference to a concept’s generic category and differentiating characteristics). On the surface, this seems reasonable. In addition, some pragmatic factors influenced the content of term banks in the early days.

For instance, it is important to remember that term banks were first conceived in the 1960s, and that prior to that, the results of terminology work had been recorded in printed glossaries (Rondeau 1984). Both print material and early computer resources had something in common: space was at a premium. Today, we have the luxury of storing enormous volumes of data, but in the early days of Terminology, this was not the case. Terminologists therefore had to be selective about the type of material that they included in terminology resources. Since the underlying goals were to standardize specialized language and to help users find the best term for their needs, there was a focus on presenting basic linguistic information in early term records, such as the preferred term, along with grammatical information (e.g., part of speech), a definition, a context, and any intra- or interlingual equivalents. The term, not the concept, was the *end product* of the terminologists’ research, and this end product is largely what was featured in term banks. This is not unreasonable. When you go to the store to buy a product, such as toothbrush, you do not typically want to purchase the component pieces

and some notes about how to assemble them into a functioning toothbrush. Most often, you are looking to acquire the end product. The process that was used by the expert to design and build the product is probably not all that important to you. You just want to buy the toothbrush – the end product – and get on with the job of brushing your teeth.

A term record is like a finished product. Translators can, in theory, consult a term bank and find the term that they need to plug into their text, and move on. However, a lot of work has gone on behind the scenes in order to build that end product. Terminologists have conducted an enormous amount of research, much of which is based on understanding the concepts and the relationships between them, and then the terminologists have distilled this knowledge into a much shorter and tidier version to record in a term record. The translator sees the end product but is not privy to all the messy intermediary stages. This would be a great situation if translation were like brushing your teeth; however, it is not. Translation is a much more complex process that involves truly understanding the concepts behind the terms. Moreover, as we have seen, a major part of understanding a concept is understanding the relations that it has with other concepts. Therefore, like terminology, translation is a knowledge-based process, rather than a strictly linguistic one. In order to be able to translate a specialized text, translators would benefit from having access to a lot of the work that terminologists have carried out in the steps leading up to the creation of the final term record (e.g., building a concept system). Sure, it might be a little messy, but it can also be instructive and informative as part of a learning process. Unfortunately, however, this information is not typically stored in a term record and so it is not accessible to the translator as part of a conventional term bank. As mentioned above, one pragmatic reason for this is that initially there was insufficient space to store everything.

In addition, in the case of concept systems, these could not be easily created or stored owing to the limitations of early technology. Often, concept systems are grasped most easily when viewed in a visual way, such as in the form of a tree diagram or a network of nodes and links. Indeed, many terminology textbooks encourage terminologists to sketch out diagrams of concept systems as part of their work (e.g., Dubuc 1978). However, early computers could process text but not graphics. There was no way to store or represent the concept systems that were often hand-drawn by terminologists. Moreover, for the terminologists who were doing the drawing, pencil and paper did not offer much support for complex multidimensional modelling. Picht and Draskau (1985), for example, had acknowledged the possibility of multidimensional classification (which they refer to as *polyhierarchies*); however, they generally discourage this practice because it risks becoming too complex, and they note that while it may be interesting in theory, it is not necessarily advantageous in practice.

Writing at the beginning of the 1990s, Sager observed that up to that point,

A great deal of attention has also been paid to the structure of conceptual systems and the best way of representing them on paper. In this area in particular, conventional thinking is still largely dominated by pen and paper processing techniques and by the relatively simple relationships that suffice for structuring documentation thesauri and respond well to human attribution and processing. Such topics, primarily of interest to theorists, have, however, found little echo in the discussion of the people directly concerned with terminology processing, notably in term banks. Here a totally different set of problems arises – not without theoretical interest – that centres on the linguistic data with which term banks have to struggle every day. [...] There is little, if any, discussion of terminological relationships nor of their representation in term banks – such relationships are not yet widely introduced in large-scale working systems. (Sager 1990, 9–10)

However, things were about to change, and Sager was one of the key influencers to usher in the era of knowledge-based terminology.

3. Knowledge-based Terminology and multidimensionality

By the 1990s, the discipline was beginning to mature, and technology was advancing rapidly. Terminologists were starting to apply terminological principles to a wider range of subject fields (not all of which were amenable to the type of hierarchical classification that works well for many physical objects). Users of terminology resources were advocating for information that went beyond that conceived mainly for linguistic standardization purposes. Meanwhile, researchers were ready to move beyond the basics, and Sager (1990) was among the first to emphasize explicitly that Terminology has not only a linguistic point of view, but also a cognitive one and a communicative one. Additionally, Sager (1990) stressed the interdisciplinary nature of Terminology, underlining that it was not merely a subfield of Applied Linguistics, but that it also stood to benefit from (and contribute to) other related fields, including Cognitive Science and Computer Science, among others. This changing landscape paved the way for more in-depth explorations of multidimensionality.

For example, one major influence from the field of Cognitive Science was the work done by Eleanor Rosch (1978) on prototype theory. The GTT subscribes largely to the classical theory of classification, which reflects the thinking of Aristotle (Felber 1984, 103) and which privileges hierarchical classification (e.g. generic-specific or part-whole relations). In the classical theory, class membership is binary: either an object belongs to a given class of objects, or it does not, making classification relatively straightforward. In addition, the classical theory

defines concepts by (individually) necessary and (jointly) sufficient characteristics. Necessary characteristics are those that every instance of a concept must have, and the set of jointly sufficient characteristics are those that no members of another class possess (and so they can be used to distinguish one class from another). Rosch (1978) presented prototype theory as a significant departure from the classical theory. Rather than defining concepts by a set of necessary and sufficient characteristics, prototype theory defines categories based on either a specific artifact of that category or by a set of entities within the category that represent a prototypical member. The prototype of a category can be understood in lay terms as being a member of a class that is most often associated with that class (e.g., a sparrow is a prototypical example of a bird, whereas a penguin is not). The prototype is the centre of the class, with all other members moving progressively further from the prototype, which leads to the gradation of category membership. In other words, not all members of the class are equally central in human cognition. In contrast to the classical view, prototypes and gradations result in an understanding of category membership not as a binary approach, but more as a network of interlocking categories which overlap.

By extension, this idea of prototypicality can apply to the characteristics of concepts too. Some characteristics may be more prototypical than others, and so they may be more commonly used as the basis for classification in a concept system. However, this does not mean that the other characteristics are wholly unimportant or that there are no instances when a classification based on other characteristics would be useful. Usually, the decision about which classifying criteria is most useful depends on the context. As noted by Bowker (1997a), there may be contexts where a group of concepts need only be classified in a single dimension to meet a specific need (e.g., in a tax schedule, vehicles may be classified only by price). However, terminologists do not have the luxury of focusing on just one dimension because the users of terminology products may have a variety of needs. Therefore, to make resources more useful, terminologists must be able to identify occurrences of multidimensionality and understand the underlying motivations for multidimensional classification. In addition, when considered collectively, the various dimensions provide a greater amount of knowledge about a concept and can lead to a more solid and holistic understanding. Inspired in part by the work of Rosch (1978) and other cognitive scientists, terminology theories (e.g., Sociocognitive Terminology as posited by Temmerman 2000, and in this volume, or Frame-based Terminology as presented by Faber in this volume, and Faber et al. 2006) began to adopt a more flexible view that allows for concepts to be organized in different ways along multiple dimensions, rather than viewing concept systems as being fixed or rigid, where each concept has a single and “correct” position within the system.

While theories from Cognitive Science provided one impetus for the movement towards Knowledge-based Terminology, another came from the field of Computer Science, and more specifically, from its subfield, Knowledge Engineering. Here, researchers were developing tools to assist with the formal structuring and representation of knowledge (Bowker and Lethbridge 1994). Knowledge engineering tools offered the possibility of generating ontologies or concept systems that incorporate the automatic inheritance of characteristics from a superordinate concept to a subordinate concept. Referring back to our earlier example of WINE, this concept can be classified into multiple dimensions according to characteristics such as “colour”, “country of origin” and “sugar content”. The dimension based on “colour” contains the concepts RED WINE and WHITE WINE. In this example, the concept RED WINE has a value of “red” for the characteristic “colour”, and this will be inherited automatically by the subordinate concepts in the hierarchy, such as CHIANTI and SHIRAZ, which are types of RED WINE. What is of particular interest for multidimensionality, however, is the fact that these knowledge engineering tools include a mechanism for *multiple* inheritance. In other words, a subordinate concept can inherit characteristics from superordinate concepts that originate in different dimensions. For example, in addition to inheriting the characteristic “red” from the superordinate concept RED WINE (in the dimension based on the characteristic “colour”), CHIANTI could also inherit a characteristic from a superordinate concept in the dimension based on “country of origin”, and it could inherit yet another characteristic from a superordinate concept in the dimension based on “sugar content”. Therefore, CHIANTI could have three different superordinate concepts – RED WINE, ITALIAN WINE and DRY WINE – and it would inherit characteristics from each. The ability to manage multiple inheritance was greatly facilitated by the mechanisms incorporated into the knowledge engineering tools, and the results of multiple inheritance could also be displayed graphically.

As noted previously, representing multidimensionality on paper was challenging for terminologists. However, knowledge engineering tools made it possible to simultaneously create and manage multiple dimensions of a concept, each of which used a different characteristic for division. The first signs of terminologists explicitly attempting to account for and address multidimensionality in terminology resources appeared in the early 1990s, when terminology researcher, Ingrid Meyer, began collaborating with Douglas Skuce, a specialist in the field of Artificial Intelligence and Knowledge Engineering. Based on the idea that users need more knowledge – not just linguistic but also conceptual or encyclopedic – Meyer and colleagues (Meyer, Bowker and Eck 1992; Meyer et al. 1992; Meyer in this volume) developed a terminological knowledge base (TKB), known as COGNITERM, which they describe as a hybrid between a conventional term bank and a knowledge

base as understood in the field of Knowledge Engineering (i.e., a highly structured set of knowledge). Multidimensionality was considered to be an important aspect of the TKB because it would allow users to gain a more holistic understanding of the concepts and would cater to a wider range of user needs and use cases, not all of which may be prototypical.

One of the main differences between a conventional term bank (e.g., TERMIUM Plus or IATE) and the TKB was that the former contained mainly linguistic information about terms, while the latter sought to include a richer variety of conceptual information. To achieve this, the COGNITERM team set out to explicitly include, encode and structure information that was often omitted from or that remained implicit in conventional term banks. One important type of conceptual information that was explicitly addressed in COGNITERM was information about conceptual relations, and this was done using graphical representations of knowledge structures.

As noted above, even though they may have been aware of the possibility of multidimensionality, terminologists had not traditionally attempted to account for it in any serious way because it was too challenging to do so using tools such as pencil and paper. As a result, information about conceptual relations had remained largely implicit in conventional term banks (e.g., in definitions or contexts). In contrast, the knowledge engineering tool allowed for an explicit graphical representation of both hierarchical and non-hierarchical relations, and it also included a masking capability that allowed users to restrict the knowledge being viewed (e.g., to show only concepts that fall within a given dimension). In this way, it addressed the complexity concern that had been raised by Picht and Draskau (1985). Although these early papers heralded the potential of TKBs for helping terminologists to manage multidimensionality, few details were provided. Indeed, Meyer et al. (1992, 959) note that multidimensionality is one of “the more troublesome methodological issues with which we are currently grappling.”

More detailed descriptions of the phenomenon of multidimensionality appeared in Bowker and Meyer (1993) and Bowker (1997a), which emphasize that it is imperative for terminologists to find a way to explicitly represent multidimensionality in terminology resources. To this end, these two works present a series of guidelines to help terminologists with two distinct tasks. On the one hand, there are guidelines for recognizing multidimensionality, and on the other hand, there are recommendations for representing multidimensionality. These two sets of guidelines, which are adapted from Bowker and Meyer (1993) and Bowker (1997a), are summarized in Table 1.

It is clear from the guidelines that working with a graphical representation is crucial for processing and presenting multidimensionality, and it has been noted above that this is a key reason that terminology researchers and practitioners had

Table 1. Guidelines for recognizing and representing multidimensionality (adapted from Bowker and Meyer 1993 and Bowker 1997a)

Guidelines for recognizing multidimensionality	Guidelines for representing multidimensionality
– Work with a graphic representation (e.g., concept map); if none exists in the available corpus, create one. – Consider all possible ways of classifying a subject field at all levels of the classification (so as not to overlook any concepts in the subject field). – Recognize signs that multidimensionality exists (e.g., use of different sets of contrasting lexical pairs, repetition of concepts at different places in a concept system). – Identify and correct false multidimensionality (i.e., when a concept is classified at one level of the hierarchy on the basis of two characteristics simultaneously that should instead be applied at <i>different</i> levels within the same dimension) (e.g. do not go from BOAT > SCHOONER but rather from BOAT > SAILBOAT > SCHOONER).	– Work on one dimension at a time, completing it as exhaustively as possible before proceeding to the next one (to minimize confusion). – Rank the dimensions in order of relative importance (e.g., based on frequency of occurrence or advice from subject experts). – Clearly distinguish the different dimensions on the graphic representation (e.g., with a label or different colours). – Clearly indicate the characteristic underlying each dimension in a way that is easily accessible to the user (e.g., in a legend). – Revise the graphic representation of the concept system as more is learned about the subject field (i.e., the concept system must be dynamic, not static).

previously neglected this often complex phenomenon. While tools available at the end of the twentieth century were an improvement over the pencil-and-paper options formerly on offer, they nonetheless had some shortcomings, such as the limitation of working in a two-dimensional space, as observed by Wright (1997, 93):

When complex multidimensional systems are represented on-screen, the primary mechanism for accommodating complexity has generally involved the use of a very large screen on which different dimensions can be viewed by scrolling through extended segments of the screen. Despite the sophistication and apparent success of such programs, it is nonetheless important to note that they are still, for the most part, working with two dimensions on the computer screen, even as earlier systems operated with two dimensions on paper.

Wright (1997) points to partial solutions used by terminologists to express the presence of multidimensionality, such as representing different dimensions in different colours, or connecting all the concepts in a given dimension with a curved line on a graph. Another possibility is to use hypertext capability to click on a given node in one concept system in order to move to another screen to display that node as part of a different concept system (i.e., a different dimension), although in this case,

users would see only one dimension at a time. In advocating for multidimensionality, Wright (1997) even suggests that, if necessary, terminologists should resort to methods such as numbered or formatted lists, rather than limiting themselves to the representation of a single dimension. Looking ahead, Wright (1997) encourages the next generation of term bank designers to explore techniques offered by object-oriented programming, finite element modelling, and CAD/CAM (computer-assisted design/computer-assisted modelling) data structures in order to utilize a full range of resources for representing complex multidimensional concept systems. As we will see in an upcoming section, successive TKB projects have attempted to implement some of these different suggestions.

4. Corpus-based Terminology and multidimensionality

Also writing in the late 1990s, when corpus-based approaches to Terminology were just beginning to emerge, Kageura (1997b, 131) suggests that corpus-based analyses constitute another way in which computer tools/resources could aid terminologists seeking to better comprehend multidimensionality. As noted above, Sager (1990) had advocated that Terminology had not only a linguistic dimension, but also a cognitive one (discussed in the previous section) and a communicative one. Corpus-based approaches examine terms in context in authentic communicative settings. Bowker (1997b) and Rogers (2004) are two researchers who conducted corpus-based investigations related to multidimensionality, and some of their main findings are summarized below.

Bowker (1997b) undertook a corpus-based study to investigate whether multidimensionality had an effect on term formation patterns. Using a series of compound terms that contained two classifying characteristics (e.g., *colour flatbed scanner*), she investigated whether the surrounding context appeared to influence the order of the elements in the term. Her findings revealed that in cases where colour capability was the main topic of discussion, the form of the term that appeared most frequently was *colour flatbed scanner*, while in those contexts where the form of the scanner (e.g., flatbed vs handheld) was the focus, then the term more often appeared in the order *flatbed colour scanner*. Based on the evidence presented in the corpus, Bowker (1997b) suggests that subject field experts are not careless or arbitrary in their use of terms but that their choice may be influenced by multidimensionality (and not univocity). In other words, an expert may consciously decide to use one term to emphasize a particular dimension at one time, and another term to focus on another dimension at another time, even though both terms are referring to the same concept. What is more, Bowker (1997b) submits

that this process is both valid and necessary in specialist communication, where there is an emphasis on precision.

Another researcher who carried out a text-based study of multidimensionality – this time from a bilingual perspective – was Rogers (2004), who undertook a textual analysis of a German-language automotive handbook from Bosch and its English translation. Rogers' work confirmed Bowker's prior general observation that different dimensions are represented through the use of different terms:

the same object – here, a material object – may be referred to in a text from various perspectives if different characteristics of the concept representing that object are activated, or if different relations to other concepts are established. These different dimensions may be said to exist not only in the text but also in the concept system resulting from an analysis of the text. (Rogers 2004, 221)

One particularly interesting aspect of Rogers' investigation is that both the German original and the English translation of the automotive handbook each included a concept system diagram. As part of the analysis, Rogers (2004) was thus able to compare the terms appearing as labels for the nodes in the concept system against the terms used in the text. What she found is that even in seemingly authoritative texts, the lexical choices of authors and translators in the text do not directly reflect the structures of concept systems as they are represented in conventional conceptual diagrams labelled by decontextualized lexical items. For instance, Rogers (2004) observed that the textualized concepts may be referred to by more generic or more specific terms in the hierarchy, or by terms that do not function as concept labels (but are motivated by a characteristic that was not selected as the classifying characteristic in the concept system). The text-based analysis led Rogers (2004) to conclude that the textual structures studied cannot be easily mapped onto conventional unidimensional concept systems for at least two reasons. Firstly, terminological synonymy and polysemy (which may be functionally motivated in a text) establish complex relations between forms and meanings, and secondly, phraseological contexts in which terms appear may vary from one language to another (e.g., for structural reasons such as grammatical accommodations in the target language, or to match stylistic conventions such as noun-oriented versus verb-oriented style). Multidimensional, rather than unidimensional, concept systems are therefore a more authentic representation of knowledge as it is reflected in text.

Meanwhile, another way that corpora have proved useful for uncovering multidimensionality is through the application of knowledge patterns (KPs) (e.g., Auger and Barrière 2008; Marshman in this volume). In brief, a KP is a lexical pattern that reveals an underlying conceptual relation. For instance, the lexical pattern *is a* often reveals a generic-specific relationship (e.g., a bicycle *is a* vehicle propelled by

pedals), while the lexical pattern *has a* often reveals a part-whole relationship (e.g., a bicycle *has a* frame, wheels, a seat and handlebars). By searching in a corpus for a term accompanied by a KP, terminologists can identify knowledge-rich contexts (Meyer 2001) that may reveal information such as conceptual characteristics that point to different dimensions for that concept.

5. Examples of terminology resources that incorporate multidimensionality

As computer capabilities continued to improve, and as our understanding of multidimensionality increased, a number of terminology projects emerged that sought to explicitly incorporate aspects of this challenging phenomenon. This section explores three such projects (PuertoTerm, BEST, and EcoLexicon), paying particular attention to the ways in which they have addressed multidimensionality. These projects are presented in roughly chronological order based on the dates of the publications that describe them. As such, it is possible to see how the different projects build on ideas from previous projects and how the resulting terminological resources become more sophisticated and comprehensive in the ways that they incorporate multidimensionality.

5.1 PuertoTerm

Writing just 15 years ago, Faber et al. (2006, 191) point to the dearth of terminological resources that incorporate multidimensionality, emphasizing “there is a great need for terminological databases constituting authentic information systems that highlight conceptual interrelatedness and underline the multidimensionality of specialized concepts.” However, these researchers also recognize that to develop such a resource, it is necessary to consider the conceptual structure of specialized domains in the context of terminology management. This is critical since the way that concepts are represented affects the way that information is configured within both individual terminology entries and the contents of each data field, particularly with regard to the description or definition of individual concepts.

Therefore, Faber et al. (2006) argue that conceptual representation needs to be more than a hierarchical list of objects that are linguistically translated into either simple or compound nouns because such hierarchies are static configurations unless conceived as part of a more dynamic structure. Specifically, Faber et al. (2006) propose the adoption of a frame-based approach, where the representation of a prototypical domain event provides a frame for the basic processes that take place within the specialized field. For these researchers, the multidimensional nature

of conceptual relations is an important element of their proposed Frame-based Terminology (FBT) approach.

In FBT, the description of specialized domains is based on the events that generally take place in them, and each area of knowledge can be said to have its own event template. As a result, part of the understanding of a specialized field can be described in terms of basic entities linked by different types of conceptual relations. Depending on the knowledge field, there is an Engineering Event, a Medical Event, and so on. Repeated actions or events are easily accommodated by the conceptual system, and therefore form the basis for conceptual event structures that are in turn used to deal with new events and actions, which are categorized within the context of such idealized frames. The construction of an event template for a specialized domain helps users to process its conceptual content more easily. The basic concepts of object, event, and attribute (characteristic) are also the basic categories for any type of terminology management, which seeks to facilitate understanding. Terminology management means configuring terms with essential information and organizing them in meaningful ways to highlight their relations to one another so that users can process them, absorb the knowledge they transmit, and remember them more clearly and easily. Flexibility is a key element in this approach given that events are dynamic and real-world entities can play different roles in them.

Faber et al. (2006) apply their ideas by building PuertoTerm, a TKB in the domain of Coastal Engineering, for which they elaborate a series of frames for the Coastal Engineering Event, including a sub-frame for the concept *TIDE*, which they describe in detail. The overall approach is corpus-based and involves the analysis of terms in context, which allows the researchers to identify different types of relations among the entities and processes that make up the frame elements. A detailed description of FBT is beyond the scope of this chapter (see Faber in this volume); however, multidimensionality is an important element that the team seeks to represent in PuertoTerm, and this is our focus here.

As part of the representation of the *TIDE* frame, the PuertoTerm developers created a schema that consists of slots and fillers. This allows the specification of the relevant information regarding, among other things, conceptual relations. A part of the *TIDE* schema that shows some of the relations is reproduced in Table 2. The items in the left-hand column indicate the conceptual relations through which *TIDE* is related to the concepts on the right. In addition to the hierarchical generic-specific relation *is_a*, there are some non-hierarchical relations, such as *caused-by* and *produces*.

Another very interesting aspect of PuertoTerm is that the researchers linked the conceptual relations directly to the definitions. According to Faber et al. (2006, 202–203), definitions have not always been accorded due importance in Terminology,

Table 2. Relations of TIDE shown in a slot-filler format (extract from Faber et al. 2006, 201)

TIDE SCHEMA: FRAME ELEMENTS TIDE	
IS_A	→ WAVE: WATER-MOVEMENT
RELATIONS	
CAUSED-BY	→ ASTRONOMICAL-BODY
AFFECTED-BY	→ ASTRONOMICAL-BODY-POSITION
PRODUCES	→ CURRENT
CHANGES	→ WATER-LEVEL
TRANSPORTS	→ SEDIMENTS
etc.	

and often, these are simply cut and pasted from other sources with little consideration for their internal structure or coherence. However, Faber et al. (2006, 203) insist that definitions are “mini-knowledge-representations, and accordingly, the organization of information encoded in definitions should be structured according to its perceptual salience”, and the definitions of other related concepts within the same category should be taken into account too. Drawing on the Functional-Lexematic Model (FLM) (e.g., Martín Mingorance 1989), the PuertoTerm team highlights that the representation of a definitional structure is similar to a conceptual representation and it may therefore show elements such as genus (i.e., referring to the domain or superordinate category to which the concept belongs), species (i.e., referring to the subdomain), differentiae (i.e., essential characteristics of a concept that make it different from others in the same category), and accident (i.e., non-prototypical characteristics of the concept). As illustrated in Figure 1, the FLM establishes the frame elements (e.g., location, material, function) in a definitional schema that directly refers to the underlying event structure of the domain, meaning that definitions are conceived as bridges between concepts and terms. This work laid the groundwork for a more robust approach to incorporating multidimensionality in flexible terminological definitions, as we will see in the work of San Martín et al. (2017), to be discussed in the upcoming section on EcoLexicon.

In their conclusion, Faber et al. (2006) assert that only through the principled analysis and design of both the micro- and macrostructure of the concepts in a specialized domain can a list of terms be transformed into an authentic information system that can aid comprehension and support knowledge acquisition.

groin hard coastal defence structure [**is-a**], located perpendicular to the shoreline [**location-of**], made of timber, concrete or rubble mound [**made-of**] in order to retard littoral drift and prevent beach erosion [**has-function**].

Figure 1. Definition for the concept GROIN in PuertoTerm (Faber et al. 2006, 205)

5.2 BEST

Fernández and Faber (2011) describe an online dictionary prototype for learners in the domain of architecture and construction, which is known as BEST (Bilingualized English-Spanish Thesaurus). As a learner dictionary, BEST has a strong pedagogical focus: users need to understand architecture and construction concepts, as well as to comprehend the relationships between concepts that represent knowledge in the domain. Fernández and Faber (2011) emphasize that the acquisition of domain knowledge depends on understanding the representation of the conceptual structures underlying the terms of a domain, and conceptual relations therefore play a very important role because they are fundamental to understanding how concepts are interconnected within a subject field. Conventional term banks tend not to contain a rich or systematically structured knowledge component, and this was something that the creators of BEST set out to address in their own learner-oriented resource. Consequently, multidimensional representation ended up being a key feature.

Taking a corpus-based approach, Fernández and Faber (2011) studied terms in context to identify their characteristics as well as their conceptual relations. Using the extended example of the concept WINDOW, which is a particularly complex concept in architecture and building construction, the authors explain how WINDOW can be subclassified in different ways according to different characteristics. In other words, all concepts in the WINDOW subdomain are specified from different perspectives or dimensions, which means that the intension of the concept (i.e., the totality of all its characteristics) is multidimensional since one characteristic can be foregrounded while others remain in the background. For example, Fernández and Faber (2011) observe that, firstly, two different types of hierarchical classification exist for WINDOW: part-whole (e.g., WINDOW has the parts FRAME, GLASS and HARDWARE) and generic-specific. Moreover, different characteristics exist at the same level of the generic-specific hierarchy, which means that different dimensions can be created based on these different characteristics. For instance, WINDOW can be classified according to the characteristic “material” (e.g., STAINED GLASS WINDOW), “mode of operation” (e.g., CASEMENT WINDOW), “function” (e.g., EMERGENCY EXIT WINDOW), or “location” (e.g., DORMER WINDOW), among others.

Although they found it essential to address multidimensionality in BEST, Fernández and Faber (2011) acknowledge that it was not always straightforward to do so, noting that while some terms had clearly defined characteristics and could be assigned to a dimension relatively easily, others presented challenges. For example, ARCHED DORMER could be placed in either the dimension based on “form” or the dimension based on “location”. The authors decided that this (and similar) entries would be included in both dimensions; however, one dimension was chosen as

the default value, even though the concept could be accessed in both. In this case, the dimension based on “form” was assigned as the default because the authors found this to be the most salient characteristic in its definition (i.e., *dormer having a semi-cylindrical roof*).

BEST goes further than the simple slot-filler display that was implemented in PuertoTerm. In BEST, conceptual maps have been created for all dimensions, and hypertext is used to represent the multidimensional concept system graphically. This allows the user to click on a given node to display textual or graphical material showing additional concept systems. In this way, layers of the multidimensional system are superimposed over one another so that users can either view a single dimension, or display one or more dimensions simultaneously. In addition, the resulting multidimensional representation is the basis for a dynamic dictionary that can be queried in different ways. However, it is worth noting that the presentation of the conceptual map did not consist of a visual network of nodes and links but was more limited (from a graphical perspective), as illustrated in Figure 2. It was largely text based, consisting of terms (representing nodes) that were grouped together using brackets, and labelled with abbreviations to represent different types of relations (e.g., HP = hierarchical partitive conceptual relation; HG = hierarchical generic conceptual relation; Non-H = non-hierarchical relation). As such, the concept map was not very user-friendly, which ultimately could have detracted from the intended pedagogical value of the approach. Note that this observation is not intended as a criticism of the efforts made by the BEST team to incorporate multidimensionality into this terminological resource, which were certainly appreciated. Rather, it is a confirmation of the challenge of doing so, and a partial explanation as to why so many terminological resources had previously shied away from taking such steps.

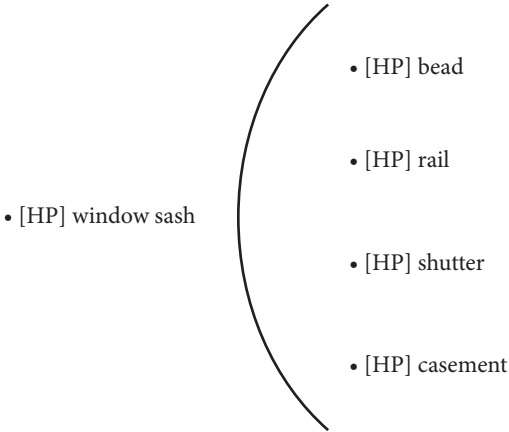


Figure 2. Extract from a conceptual map for the concept WINDOW taken from BEST (Fernández and Faber 2011, 14)

5.3 EcoLexicon

In this final example, León-Araúz, Reimerink and Faber (2013) describe EcoLexicon, a multilingual TKB on the environment that evolved from PuertoTerm and that supports users who need to understand or produce specialized texts in this domain. As expressed by its creators, “The main and ultimate goal of EcoLexicon is user knowledge acquisition, which can only be achieved if TKBs account for the natural dynamism of knowledge mainly caused by context and multidimensionality” (León-Araúz, Reimerink and Faber 2013, 144). EcoLexicon has a visual interface with different modules for conceptual, linguistic, and graphical data. In addition, while the previously discussed PuertoTerm and BEST projects were corpus-based to some degree, the EcoLexicon project uses corpus-based techniques more extensively, including the use of KPs, which are lexical instantiations of underlying conceptual relations.

Taking an iterative approach, the team uses well-known KPs that represent both hierarchical and non-hierarchical relations to search the corpus (e.g., *is a*, *part of*, *delimited by*, *causes*, *located in*, *affected by*, *made of*, *has function*, *result of*, *takes place in*, *affects*, *phase of*, *attribute of*), which can reveal instances of multidimensionality. For instance, the concept EROSION can be classified according to the characteristic “direction” producing a dimension with the subconcepts LATERAL EROSION and HEADWARD EROSION. Additionally, EROSION can be classified according to “result”, producing a dimension with subconcepts such as SHEET, RILL, and GULLY. Other dimensions for this concept exist also.

Although multidimensionality is ideally seen as a way to enrich traditional static representations of knowledge by enhancing knowledge acquisition through different points of view in the same concept system, León-Araúz, Reimerink and Faber (2013) discovered (as Picht and Draskau [1985] had predicted decades earlier) that this can lead to information overload, which can actually end up jeopardizing knowledge acquisition. They give the example of WATER, which is a particularly versatile concept for which 72 different conceptual relations for the first level of the hierarchy were identified in the corpus. The resulting conceptual graph for this concept was so complex as to be virtually unusable. Indeed, user testing of EcoLexicon by López Rodríguez, Buendía Castro and García Aragón (2012, 71) confirmed that many users found the dynamic concept systems to be “confusing whenever the concept had many relations with other concepts, because they were displayed simultaneously.” User suggestions to address this concern included displaying only the most basic relations by default, or applying colours or different fonts to allow better discrimination of different concepts and relations.

The EcoLexicon team have continued to work to improve the TKB, and as reported by San Martín et al. (2017), steps taken address the issue of information

overload resulting from multidimensionality include: (i) allowing users to filter networks by relation type; (ii) providing different access modes to the concept visualization (network mode, tree mode, path mode); and (iii) offering a recontextualized view of concepts according to domain-based contextual constraints. With regard to filtering, three categories of conceptual relations are identified in EcoLexicon (generic-specific, part-whole, and non-hierarchical) and each is displayed in a different colour. Users may choose to display/hide one or more of the relation types. Meanwhile, a network mode display can show all types of relations in the network. In contrast, the tree mode generates a generic-specific hierarchy for the concept, while in path mode, users select two concepts to represent the beginning and end of a path and the tool calculates and displays the shortest distance between them.

Moreover, users may opt to generate a recontextualized network by choosing one of the contextual domains from a drop-down menu. Recontextualization is an approach developed by León-Araúz, Reimerink and Faber (2013) to address the problem of information overload while enhancing the representation of multidimensionality. According to León-Araúz, Reimerink and Faber (2013), any specialized domain contains subdomains in which particular dimensions are more or less important depending on the activation of specific contexts. Recontextualized networks are reshaped depending on how the relational behaviour of concepts varies with perspective. Instead of representing all possible domains of a concept, conceptual propositions are activated or suppressed according to their salience in given subject fields. Recontextualization is thus based on prototypes and context.

Another key addition to EcoLexicon that incorporates multidimensionality is the adoption of a flexible terminological definition approach (FTDA) (San Martín et al. 2017). In line with the recontextualization of conceptual networks as described above, the FTDA applies thematic contextual constraints (implemented in the form of subdomains, such as Hydrology, Chemistry, Energy, and Engineering). More specifically, it consists of a set of definitions of the same concept made up of a general definition (in this case, one encompassing the environmental domain as a whole) as well as additional definitions describing the concept from the perspective of the subdomains in which it is relevant. The representation of the characteristics activated by concepts in accordance with the context allows a definition to better meet the needs of a specific user. This also entails the inclusion of prototypical characteristics in the definition (i.e., characteristics that are not applicable to the concept in every context, but which are relevant in a given context).

6. Concluding remarks and future directions

Over time, the notion of ‘concept system’ has continued to hold special significance in Terminology; however, our thinking about how these systems look and operate has evolved. Rather than continuing to view concept systems as being fixed or rigid, where each concept has a single and so-called correct position within the system, terminology theories have begun to adopt a more flexible view that allows for concepts to be organized in different ways along multiple dimensions. Overall, this has the potential to lead to a more holistic understanding of a concept, and it may make a terminological resource useful to a wider range of users.

Successive attempts to integrate multidimensionality into terminological resources have revealed challenges, particularly in regard to the potential for information overload, but they have also highlighted the creativity of researchers for finding solutions. In terms of work remaining to be done, there is plenty. For instance, although the EcoLexicon team has focused on subdomains as a basis for generating narrower and recontextualized networks, there may be other options for meaningfully reducing information overload relating to multidimensionality. Rogers’ (2004) bilingual study revealed that even in a technical domain such as automotive engineering, language and culture can influence multidimensionality. Since most explorations of multidimensionality to date have focused on scientific and technical domains, it would be valuable to explore this phenomenon in more detail in more humanities-oriented domains, to see if culture plays a more significant role. Continued work on knowledge patterns may lead to insights as to which of these appear most productive for investigating multidimensionality in corpora. Finally, continued and rigorous user testing of terminology resources that implement multidimensionality will also be informative for pointing to new directions for filling gaps and continued improvement.

Terminology and ontologies

Elena Montiel-Ponsoda

ETSI Informáticos – Universidad Politécnica de Madrid

This chapter revisits the relation between terminologies and ontologies, focusing on those aspects that lie at the intersection of the two worlds. Terminological resources, methods and tools that follow the conceptual paradigm are analyzed to explain how close they are to ontologies. Then, current methodologies in Ontology Engineering are reviewed from a terminological perspective to highlight those aspects of terminological practice that are becoming increasingly relevant in an ontology development process. Finally, current trends and models in Ontology Engineering whose aim is to enrich ontologies with linguistic (terminological) descriptions are examined in the light of their use in Artificial Intelligence and Natural Language Processing tools, pointing to a renewed convergence of these two worlds.

Keywords: Ontology Engineering, Concept-Based Terminology, terminological knowledge base, termontography, Artificial Intelligence, Semantic Web, NLP applications, Ontolex-lemon, linked data

1. Introduction

Ontologies and terminologies are the object of study of Ontology Engineering and Terminology, two disciplines that have many aspects in common. Firstly, what lies at the core of both is communication, since both facilitate communication between different interlocutors. In the case of ontologies, interlocutors can be two software agents (computers), or a software agent and a human. The main interlocutors in Terminology are humans, specifically domain experts, sometimes with different levels of expertise. Because of differences in users and uses, the resulting products provide different types of information. The information is also structured, formalized, and exploited differently.

Ontologies represent and structure knowledge in a way that can be processed by machines, so that reasoning can be automated and information can be inferred. This generally results in simplified knowledge models with scarce linguistic data.

Depending on the aim of the resource and theoretical stance, terminologies range from multilingual lists of equivalent terms to resources that try to account for the syntagmatic and paradigmatic relations of terminological units, or others that have combined a knowledge base with a term base.

Despite the differences that ontologies and terminologies exhibit, there have been various attempts to bring these two worlds together and make the most of the similarities they share. Especially in the areas of Computational Linguistics, Natural Language Processing (NLP), and Artificial Intelligence (AI), the reuse and integration of curated language (terminological) resources in computer applications that process language (e.g., machine translation, text summarisation or virtual assistants) have provided major benefits.

This chapter explores the relation between ontologies and terminologies from a terminological viewpoint. Section 2 defines ontologies and explains their background. Section 3 gives an overview of terminology work, resources and tools that have relied on knowledge representation with a view to explaining how close they are to ontologies, and how they could be potentially converted to Semantic Web formats and used in ontology-based NLP applications. Section 4 explores those activities in the ontology development process that involve terminological analysis. Section 5 discusses how ontologies can be enriched with linguistic descriptions to further bridge the gap between the conceptualization formalized in ontologies and the semantics of language. Finally, Section 6 summarizes the conclusions that can be derived from this research.

2. The quest for knowledge representation

The interest in describing and explaining “the first causes (*aitia*) and the principles (*archai*) of things” finds its roots in what is known as the “Metaphysics” treatise by Aristotle. As summarized in Cohen and Reeve (2020, 2), Aristotle proposed “the study of being *qua* being”, i.e., the study of beings or entities (*onta*), “of things that can be said to be”. His definition distinguished the philosophical approach to the study of entities from Natural Sciences or Mathematics, which studied entities under different aspects (such as changing things subject to the law of nature or as countable and measurable things, respectively).

As reported in Cohen and Reeve (2020, 4) Aristotle proposed an inventory of all the possible things that exist (i.e., substance, quality, quantity, relation, place, time, position, state, action, and affection). Of the items on the list, ‘substance’ holds a privileged position.

Substances are unique in being independent things; the items in the other categories all depend somehow on substances. That is, qualities are the qualities of substances; quantities are the amounts and sizes that substances come in; relations are the way substances stand to one another.

Substances are entities, and the rest of the categories depend on or refer to these entities. In addition to this relation of dependence, Aristotle pointed out the relation of inheritance between substances, especially between universals (general concepts) and particulars (or individuals), which resulted in a tree structure.

Such ideas have engaged philosophers throughout the centuries, and although many different groupings and classifications of categories have been proposed, what remains is the basic notion of entities, described by common properties and by relations, as in the knowledge representation systems represented by ontologies, amongst others.

2.1 Ontologies from the computational perspective

The idea of categorization and structuring is at the core of traditional classification schemes (taxonomies in Natural Sciences, thesauri in Library and Information Science) and *computational ontologies* (Guarino, Oberle and Staab 2009, 1). Artificial Intelligence (AI) researchers used categorization to build ontologies in the sense of computational artifacts for knowledge exchange in AI systems. Consequently, ontologies are the result of a knowledge engineering activity that computationally represents knowledge to enable agents on the Web to automatically reason about that knowledge and interoperate (i.e., exchange information). Gruber (1993, 11) writes:

Pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among agents. Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner.

In addition to agreeing to the same vocabulary or universe of discourse, agreement is to be reached on the representation language format and on the communication protocols. It is only then that knowledge-level communication can succeed (Gruber 1993, 3). This justifies the interest in this area for standardization activities that propose interchange protocols, formats, and models that are to be adopted by practitioners to guarantee the success of communication between machines. The organization in charge of the development of standards for the Web as an information space is the World Wide Web Consortium, better known as W3C (2020), led by Sir Tim Berners-Lee, one of the inventors of the Web.

The 90s was the decade that witnessed the birth of the first computational ontologies. Gruber (1993, 3) defined ontology as “an explicit specification of a conceptualization”. By ‘conceptualization’ he understood “an abstract, simplified view of the world that we wish to represent for some purpose” (Guarino, Oberle and Staab 2009, 3); by ‘explicit’, that conceptual descriptions and relations are to be clearly stated. Borst (1997, 12) added the notion of ‘shared conceptualization’, which means that a useful ontology expresses a generally accepted view of a community, and not only that of an individual. Studer, Benjamins and Fensel (1998, 185) expanded this definition as follows:

Conceptualization refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine-readable. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private of some individual, but accepted group.

Ontologies are commonly divided into domain ontologies and fundamental or upper-level ontologies. Domain ontologies are contextualized in or restricted to a certain domain or area of knowledge. Upper-level ontologies organize “general concepts that are common across the domains and give general notions, which all the terms in existing ontologies should be linked to” (Gómez-Pérez, Fernández-López and Corcho 2003, 71). Traditional examples of the former are the Gene Ontology or the PRotein Ontology (PRO) in the medical domain (both accessible at The OBO Foundry 2020). Examples of upper-level ontologies include SUMO (2020) or Mikrokosmos (O’Hara, Mahesh and Niremburg 1998).

2.2 Ontologies as artifacts that interact with machines and humans

Regardless of type, the notion of ontology raises a difficult and recurrent question: Can a certain epistemological community agree on one (and only one) view of the world? The answer is probably negative, but they may want to agree on a certain interpretation of the world for certain purposes. As Uschold (1998, 12) claims:

An ontology may take a variety of forms, but necessarily it will include a *vocabulary of terms*, and some *specification of their meaning*. This includes definition and an indication of how concepts are inter-related, which collectively impose a structure on the domain and constrain the possible interpretation of terms.

Therefore, if a community is willing to accept a certain interpretation of the terms represented in the ontology, they may agree to the use of the ontology in an application scenario, even if it means making some concessions or oversimplifying the

description of the domain. This would alleviate the burden related to communication between systems, but would probably miss communication between systems and humans.

If humans are also to play the role of interlocutors in this communication act, then the vocabulary of terms and the specification of their meaning should reflect in as much as possible the meaning of terms as they occur in specialized discourse. This may seem to contradict the simplified view of the world that ontologies inevitably represent, but is in line with research efforts that propose ways to complement ontologies with rich linguistic descriptions (see Section 5).

Some of the most common application scenarios in which ontologies interact with human language are semantic annotation, and the search and retrieval of information (Bürger and Simperl 2008). When used in search-based applications, the domain knowledge captured in the ontology can enhance keyword-based search algorithms by taking advantage of the subsumption relations or logical constraints formalized in the ontology. For example, when searching for ‘collective dismissal’, it might be useful to retrieve documents that contain the more general terms, ‘termination of employment’ or ‘dismissal’, or the variant ‘collective redundancy’, while avoiding documents that use the term, ‘unfair dismissal’. In these scenarios, an ontology that captures the real use of terms in discourse, i.e., a linguistically grounded ontology, will enhance performance.

2.3 Components of a domain ontology

Broadly speaking, ontologies structure knowledge by doing the following: (i) identifying and naming the set of concepts that describe a domain; (ii) listing the main properties of those concepts; (iii) establishing relations between concepts. In OWL (2012), the most popular ontology language of the Semantic Web,¹ *classes* (`Class`) identify specific or abstracts concepts; and *properties* are divided into *data type properties* (`datatypeProperty`) and *object properties* (`objectProperty`). Accordingly, data type properties refer to features or characteristics that define concepts, and object properties represent dependencies between concepts, or how concepts relate to each other (taxonomic, partitive, *ad hoc* relations). The objects that belong to a certain class are *individuals*, and there is a specific type of property called *axioms*, namely, restrictions imposed on classes or relations that are taken into account when reasoning about the knowledge that is expressed by the ontology. Examples of axioms are when two classes are disjoint and exhaustive, i.e.,

1. Term coined by Tim Berners-Lee in 1998, (see also Berners-Lee, Hendler and Lassila 2001). Nowadays, the term *Web of Data* has become popular and is used as a synonym of *Semantic Web*.

the individuals of one class cannot belong to the sibling class, and the two classes together account for all individuals that exist.

The syntax underlying OWL is called RDF, Resource Description Framework, (RDF 2020), and it provides a way to structure ontology elements into statements or triples of *class-predicate-object* (e.g., *Class-dataTypeProperty-value* or *Class-objectProperty-Class*).

This results in a graph structure of nodes connected to other nodes through edges (see Figure 1). This is a key aspect of the standardized languages or formats used in the Semantic Web, because they allow users to link any piece of data in as much as hyperlinks link documents. And such a common data representation is the fundamental value and differentiating aspect of the Semantic Web.

To make ontologies understandable to human users, ontology elements are usually accompanied by simple labels or designations in natural language (*label*). As its name suggests, a label provides “a human-readable version of a resource’s name” (RDF 2020). This property is accompanied by another property called *comment* (*comment*). A comment “may be used to provide a human-readable description of a resource” (*ibid*). If multilingual labels are to be associated with the same ontology class, several labels can be used. Labels try to reflect the vocabulary used by a certain community when describing the knowledge modeled in the ontology. However, ontology labels may not capture the linguistic knowledge necessary for NLP applications that rely on ontologies.

To illustrate what has been explained so far, we refer to the *GoodRelations* ontology (Hepp 2008, 329–346). *GoodRelations* is an ontology of e-commerce currently used by Google and Yahoo. In this ontology, *MobilePhone* or *TVSet* are types or subclasses of the *ProductOrService* class, whereas *category* or *color* are data-type properties describing the type of product. Object properties or relations are established with *BusinessEntity* or *DeliveryMethod*. *UPS* and *MRW* are individuals of the class *DeliveryMethod*; and an axiom could be imposed on delivery methods, stating that some manufacturers only deliver in pick up mode. When arranged in triples, this results in a graph, as depicted in Figure 1 (the nodes of the graph are ovals; edges are labeled arrows; and rectangles represent instances). Should an online shop wish to rely on such an ontology to publish its catalog of products, a search engine would be able to answer user queries like “Show me only those products of category *TVSets*, manufactured by *SONY*, which can be delivered home by *UPS*”. The triple that states that a product or service can be delivered by a certain delivery method is formalized as follows:

ProductOrService-availableDeliveryMethod-DeliveryMethod.

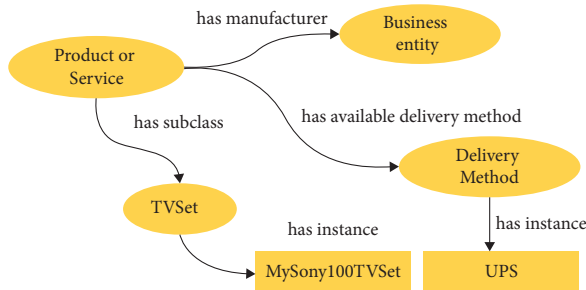


Figure 1. Triples of a simplified e-commerce ontology

Nowadays, ontologies are the building blocks of the Linked Data initiative (Bizer, Heath and Berners-Lee 2009), which is an effort to publish and connect structured data on the Web coming from multiple data sources. They also provide the framework for the construction of Knowledge Graphs (Pan et al. 2017) to harmonize and connect data in large organizations.

3. Compiling terminologies from a conceptual perspective

The idea of having an interlingua to facilitate communication is very appealing, and has also been pursued in other areas. If we replace software agents by humans, and refer to knowledge exchange or communication between experts, we find abundant parallelisms with traditional approaches to Terminology. In fact, this was the first and foremost objective of Wüster (1979) when proposing the use of terms to label concepts in knowledge structures, assuming that they should be defined in an unambiguous manner to guarantee univocity in expert-to-expert communication. This approach is at the core of what is known as the General Theory of Terminology (GTT) and has influenced some trends of terminology work and Ontology Engineering (see chapters by Humbley and Candel in this volume).

3.1 The conceptual paradigm applied to term banks

One of the projects in which Wüster was involved was the compilation of the first edition of the *International Electrotechnical Vocabulary*, whose latest version is available online (International Electrotechnical Commission 2020). This resource, which is titled *Electropedia*, is a controlled vocabulary that contains more than 22,000 terms in English and French, with translations in more than 15 languages.

The methods followed in its creation are the ones proposed by ISO for Terminology work (ISO/CD 704 2009) (see also the chapter by Wright in this volume). By adopting these methods, experts observe and examine objects and their properties, and categorize them into abstract classes. Even if no relations are established between the terms in *Electropedia* (other than translation equivalents), the methods proposed by ISO stress the importance of developing a concept system based on the relations between concepts, since it helps in delimiting the intension and extension of concepts. The main types of relations considered are hierarchical (generic and partitive), as in traditional taxonomies and ontologies, although some associative relations are also considered.

In this regard, the conceptual structure proposed by ISO for the creation of terminological resources would match the one adopted by most ontological resources. However, no support is provided for the modeling of conceptual relations beyond the graphic representations of generic and partitive relations based on UML (ISO 24156-1 2014). ISO proposes a standard for the formal representation of computerized terminology databases, namely, the Terminological Markup Framework or TMF (ISO 16642 2017) (and the associated TermBase eXchange format or TBX). This standard reflects the common structure of traditional concept-based term banks such as *Electropedia*, but fails to represent complex conceptual structures.

As for the methods to follow, the ISO principles for terminology work suggest a top-down approach, assuming a gathering of authoritative experts that agree on the characteristics of concepts, on how to define them, and what relations to include for the creation of a meaningful concept system. As stated in the 2020 version of (ISO/CD 704), currently under revision, this is a convenient approach when dealing with ‘material objects’:

If possible, terminological analysis should start with those concepts corresponding to material objects, since the characteristics are more easily abstracted given that the properties of those objects can be physically observed or examined.

Such a realism-based approach has also been popular amongst modern philosophers involved in the construction of ontologies in the biomedical domain (Smith 2004; Ceusters and Smith 2006). As claimed in Smith (2004, 79), good ontologies are devoted precisely to the representation of entities as they exist in reality. Only those terms that correspond to *universals*, and thereby also to instances, can be considered concepts in ontologies. These authors avoid the use of the term *concept* in favor of *universal* because they understand concepts as subjective entities used by a community of language users.

Even when acknowledging that this top-down approach is the most appropriate one for some areas, this view may fail to account for some aspects of real communication between specialists (Cabr  1999b, 129). For instance, the role that language

plays in the conception and communication of concepts, and the fact that many concepts cannot be so clearly delineated (the content of a term is never absolute), but relative to context and domain (Cabr  1999b, 132). In this sense, the principles and methods in ISO/CD 704 also foresee a bottom-up approach for ‘immaterial objects’ in which documents can be used as the starting point of terminology work. However, polysemy and synonymy, which are so relevant in specialized languages,² are to be avoided in favor of harmonization:

Incidences of synonymy, polysemy or homonymy usually lead to the need for concept harmonization or term harmonization, which is part of the standardization process. (...) To reduce duplication and to reduce the cost of terminology work, efforts should be made to harmonize whenever minor differences exist.

All in all, term banks have proven useful as controlled vocabularies for indexing tasks or to support technical writing in large organizations. A similar goal has been followed by some ontologies that act as controlled vocabularies to classify (and retrieve) information items in a repository. In these scenarios, both controlled vocabularies and ontologies are used as a means to formalize the kind of things that can be talked about in a system or context (B rger and Simperl 2008, 586). However, by adopting such a prescriptive approach, these resources may miss the real use of terms in specialized texts. This limits its use by NLP applications that need to interact with documents, which is why descriptive approaches are also common in Ontology Engineering. Additionally, when relying on certain representation formalisms, they may fall short of providing a flexible and comprehensive conceptualization of a domain in which relations are established between concepts.

3.2 Term banks for translators

A similar approach to the one followed in Electropedia, but with other users and uses in mind, has been traditionally taken by term banks in multilingual governmental organizations. Two well-known examples of concept-based term banks are *TERMIUM Plus*[®] (2020), the term bank of the Canadian Federal Government Translation Bureau, and IATE, the InterActive Terminology for Europe, which supports translators at the European Union institutions (Zorrilla-Agut and Fontenelle 2019). The term records in such resources are assigned to a specific domain and contain information items mainly intended to describe the concept and offer solutions to translation problems.

2. As stated in Temmerman (2000, 133), polysemy is the consequence of changes in conception over a period of time, and synonymy reflects different perspectives in specialized discourse.

As claimed in L'Homme (2006a, 58), most term records “are prescriptive in the sense that one term is preferred if a concept has several different names”. This is very much in line with the designations or labels associated with ontology classes (see Section 2). Indeed, the conceptual nature of these resources has enabled the conversion of term banks into models that rely on Semantic Web standards. This is the case of a subset of IATE term records, freely available for download, as described in (Cimiano et al. 2015). In this work, the authors converted the term records available in TBX format into RDF (RDF 2020). Terms records were translated into concepts and referred to utilizing the term record number (e.g., IATE_84). The rest of information items, namely, subject field, term type, reliability code, part-of-speech, and grammatical number were associated with the concept as object and data-type properties. Finally, term records were grouped by language, and translation equivalents were established at the concept level (by pointing to the same term record).

It goes without saying that the work of terminologists and translators in the creation of these resources has played (and continues to play) a crucial role in their respective institutions and beyond. From a computational perspective, and considering that the conceptual structure of these resources makes them particularly suitable for their conversion into semantic objects, they could be easily used in combination with domain ontologies, or integrated and reutilized in NLP applications, though, not to their full potential. Traditional multilingual term banks assume that human translators analyze the descriptions in term records (definitions, notes, language usage observations) and infer relevant information regarding the semantic, syntactic, and pragmatic behavior of terms. This information is implicit in term descriptions. Moreover, conceptual relationships between terms are neither documented nor made explicit. Therefore, some information extraction work must be done prior to bringing these resources to their full potential in the Semantic Web.

3.3 From term banks to terminological knowledge bases

With a broader perspective that includes expert communication as well as document retrieval, some computer applications went a step further and integrated term descriptions in conceptual representations. One of the first projects in this regard was COGNITERM (see chapter by Meyer in this volume). COGNITERM (Meyer, Bowker and Eck 1992, 159) was defined as a hybrid between a conventional term bank and a knowledge base. It aimed at explicitly representing the relations between concepts and terms, using a corpus as a starting point.

In comparison to traditional term banks, this terminological knowledge base provided information in a far more structured manner by representing term relations and providing the definition of the concept as a template in a characteristic-value format with previously defined characteristic types. This application also included a

graphical representation of the subject field and accounted for multidimensionality since it offered different views of the graph, depending on the dimension selected. This kind of graphical representation is common in current terminology management and ontology editing tools.

Another terminology management system that brought together the ontology and the terminology worlds was OntoTerm[®] (Moreno Ortiz 2000). The system consisted of two fundamental modules: an ontology editor and a termbase editor. The starting point in the creation of the terminological knowledge base was a core ontology of 21 basic concepts from the upper-level Mikrokosmos ontology, which was specified for the domain at hand. Ontology concepts were created prior to the language-specific terms, and data categories (according to ISO 12620 2019) were used to describe the terms in consonance with traditional terminological practices. As in the case of COGNITERM, this system also relied on a predefined set of category templates, which ‘dissect’ the set of properties that define the concepts (somehow equivalent to data-type and object properties in ontologies). Systematically, all subcategories belonging to an upper-level category were defined according to the pre-defined templates. It is worth mentioning that OntoTerm[®] allowed users to make explicit references to the documents from which the terms had been obtained, which added a third module to the system, namely, a bibliographical module.

Various terminological knowledge bases in the medical domain were implemented with this system: GENOMA-KB (Cabr   et al. 2004) and OncoTerm (L  pez-Rodr  guez et al. 2006). As described in L  pez-Rodr  guez et al. (2006), apart from documents, lexicographic definitions found in dictionaries were used to understand the conceptual structure underlying the domain and to classify terms according to hierarchical relations.

This new wave of terminological resources was decisive in two ways. Firstly, it demonstrated the importance of domain-specific corpora in a terminological endeavor (Cabr   2006). Secondly, it initiated a research line focusing on the use of computer applications at different stages of the terminology development process. This signified “moving away from conventional term records and closer to semantic networks as used in AI” (Barri  re 2006, 81), to combine ontologies, terminological data, and bibliographic information. The methods followed in the compilation of these resources share significant commonalities with current practices in ontology development, especially in the acquisition of domain knowledge, as will be outlined in Section 4.

One of the most recent evolutions of such a terminological knowledge base has been EcoLexicon (Faber, Le  n-Ara  z and Reimerink 2011; Faber and Buend  a-Castro 2014), a multilingual resource in Environmental Science. In its development, the theoretical tenets of Frame-Based Terminology (Faber et al. 2006; Faber 2015, this volume) were applied. As for the conceptual representation of terms, this

theory suggests the use of a linguistically-based ontology, i.e., a conceptualization “based on information extracted from specialized texts and the structure or terminological definitions” (Faber 2015, 19), and does not recommend that experts be involved in the initial knowledge extraction phase, but rather in the subsequent validation of the specialized knowledge retrieved.

Since 2003, this resource has been enlarged and updated to become a compulsory “one-stop shop” for language professionals in this field. It is now integrated in a CAT tool (EcoLexiCat) and complemented by corpus and concordance tools (León-Araúz, Reimerink and Faber 2019). This is an outstanding example of a resource that combines a knowledge representation in an ontological fashion with the linguistic descriptions demanded by language professionals. The information in this resource can be freely accessed through a user interface and is in line with the most current trends on data integration. Up to now, and to the best of our knowledge, it has not been implemented in a standard formal language, but this is currently ongoing work (San Martín et al. 2020).

These projects also evidence that certain areas of knowledge, especially in the Life Sciences, seem to lend themselves more naturally to a conceptual approach. In fact, Biology and Biomedicine are the areas that have most successfully adopted ontologies.³ From a multilingual perspective, obtaining equivalents in other languages appears to be almost immediate, as these areas tend to be more internationalized than others that are more culturally rooted (Law or Geography). As for GENOMA-KB and OncoTerm they are no longer available online. As so often occurs with many valuable resources, once the project is over, they can no longer be maintained. If they were to be reused in ontology-based NLP applications, it would be necessary to convert them to current Semantic Web formats. However, this could be a relatively straightforward task in the light of the similarities of these resources to ontologies.

3.4 Termonological databases: Bringing together Terminology and ontology practice

The multilingual perspective was the main challenge of a project in which an ontology and a multilingual terminological database were combined in a *termontological* resource, within the context of the Terminography approach (Temmerman 2000, chapter in this volume). As happens with multilingual termbases in translation memory tools, it was believed that a resource that would allow translators

3. Unified Medical Language System (UMLS), OBO Foundry, BioPortal or Bio2RDF are good examples of ontologies, ontology repositories, and related initiatives in this domain.

to understand the conceptual structure underlying those terms (using the links between concepts) would meet the needs of these and other language professionals.

As explained in Temmerman and Kerremans (2003), a shift towards the use of ontologies as knowledge repositories in terminography⁴ was the natural evolution of terminological knowledge bases from a sociocognitive approach. The authors suggest combining top-down and bottom-up approaches in a sort of *middle-out approach* to capture and represent both domain expertise in collaboration with field specialists, and the information contained in a domain corpus. The result of the top-down approach is an initial framework of categories. The result of the bottom-up one is knowledge retrieved from textual material reflecting culture-specific categorizations. By combining the two, a terminological database is created that can be directly exported as an ontology.

Interestingly enough, in this approach, domain-experts play an essential role in the initial stages of the terminography work, namely, when compiling the domain-specific corpus and when agreeing on the main classes to be represented in the resource. As will be explained in Section 4, very similar tasks are performed in the initial tasks of an ontology development process.

In the multilingual project outlined in Temmerman and Kerremans (2003), the role of the terminographer is to lexicalize all ontology elements that participate in a *subject-predicate-object* triple in several European languages. Those lexicalizations can correspond to real terms in natural language or to artificially created ones (Kerremans and Temmerman 2004). The main challenge for terminographers was that they encountered terms in the corpus with no correspondence in the ontology, either because the concepts represented in the ontology had been coined in English texts, or because they did not have the same granularity level in all languages. Taking this into account the authors decided to create an intermediate categorization framework where culture-independent categories were represented as well as a set of intercategorical relationships to map culture-dependent categories and terms.

This terminological approach to multilingual *termontologies* claims that a culture-independent task-oriented categorization is needed to accommodate the culture-specific terms. This is a common issue in knowledge areas without international consensual representations (contrary to what is usually the case in Life Sciences). These areas are those that deal with culturally bounded domains, such as Law. There is no valid, commonly shared representation that can accommodate cultural divergences. In this regard, Montiel-Ponsoda (2011) analyzed various projects involving multilingual ontologies, and concluded that it was necessary to distinguish between internationalized domains and culturally-influenced domains.

4. In this paper *termontography* is presented as a newly coined term by the authors to stress the combination of terminographical methods with Ontology Engineering methods.

The former rely on one shared conceptualization and a multilingual layer, whereas the latter have to necessarily account for cultural divergences. This means either resorting to an intermediate conceptualization (or *interlingua*, as in the EuroWordNet project, see Vossen 2002), or using a language (terminological) layer powerful enough to make divergences explicit. This is the purpose of some models that have been proposed in the ontology modeling literature.

A different theoretical stance that also combines an ontology with a termbase, but is based on the premise that “a concept does not need a term to exist” is Ontoterminology (Roche, 2015). In this approach, formal ontologies are conceptualized by experts and represented in a formal language before ‘the right words’ to talk about the concepts are found. According to Roche (2015, 128–133), if ontologies are derived from scientific discourse, then they cannot be considered formal ontologies, but rather ‘textual ontologies’, and this would “reduce terminology to a specialized lexicography”. In this approach, conceptualizations extra-linguistically model a reality and are standardized (in a formal, logical language). It is also necessary to reach a consensus regarding the definitions and concept identifiers. Only then can terms and term definitions be accounted for and mapped onto the ontology concepts. This work is also supported by a software environment (TEDI-OntoTerminologyEDItor)⁵ that can be freely downloaded from the Ontoterminology website and offers export options to most popular ontology languages.

The resulting resources are probably not designed for language professionals, but rather for use in standardization processes, as was the case of Electropedia. However, agreeing on a shared conceptualization that extra-linguistically models reality in a top-down fashion is not without difficulties. As discussed in Section 1, some commitment from ontology stakeholders is needed, but ontologies have to somehow account for the language and the semantics that real users understand, if they are to be exploited in tasks that involve human interaction.

Although ontologies tend to be seen as the ideal resources from the perspective of the GTT, the next section explains the best practices in ontology development used by ontology engineers for knowledge elicitation from texts. Moreover, as has been already mentioned, one of the main uses of ontologies is to annotate documents to support information retrieval tasks. The richer the linguistic descriptions associated with concepts, the better. As a result, recent years have witnessed a growing interest in the lexicalization of ontologies, that is, in verbalizing ontology concepts and properties with richer linguistic descriptions in an attempt to bring lexical semantics and ontologies closer.

5. TEDI can be accessed at the following URL: <http://ontoterminology.com/tedi>.

4. Terminology practices in Ontology Engineering

Ontology Engineering is defined as “the set of activities that concern the ontology development process, the ontology life cycle, the methods and methodologies for building ontologies, and the tool suites and languages that support them” (Gómez-Pérez, Fernández-López, Corcho 2003, 5). With the aim of supporting developers in building ontologies, several guidelines, best practices and self-styled methodologies⁶ have been proposed, usually accompanied by ontology development environments or tools. Figure 2 illustrates the main activities, actors, and expected results, as proposed by the recently published Linked Open Terms (LOT) methodology (Poveda-Villalón, Fernández Izquierdo and García Castro 2019).

As shown in Figure 2, this methodology identifies four main activities in the ontology development process in which ontology developers are expected to interact with domain experts and final users: (i) ontology requirements specification; (ii) ontology implementation; (iii) ontology publication; and (iv) ontology maintenance. To assess the role that terminology practice could play (and are actually playing) in the development of ontologies, and, as a result, in the consolidation of the Semantic Web, we analyze the main steps involved in activities (i) and (ii) from a terminological perspective.

4.1 The role of terminology analysis in knowledge acquisition

The Ontology requirements specification activity results in the creation of the ORSD, the Ontology Requirements Specification Document. The ORSD summarises the objective of the ontology, its uses and users, and the functional requirements that it should fulfill, i.e., the particular knowledge to be represented by the ontology.

Most, if not all, of the proposed methodologies suggest using the so-called *Competency Questions (CQ) activity*⁷ to elicit that knowledge. With this purpose, ontology developers assist domain experts in formulating questions (or statements) that the ontology will be able to answer once implemented. Examples of the CQs from the European SEEMP project for public e-employment services are

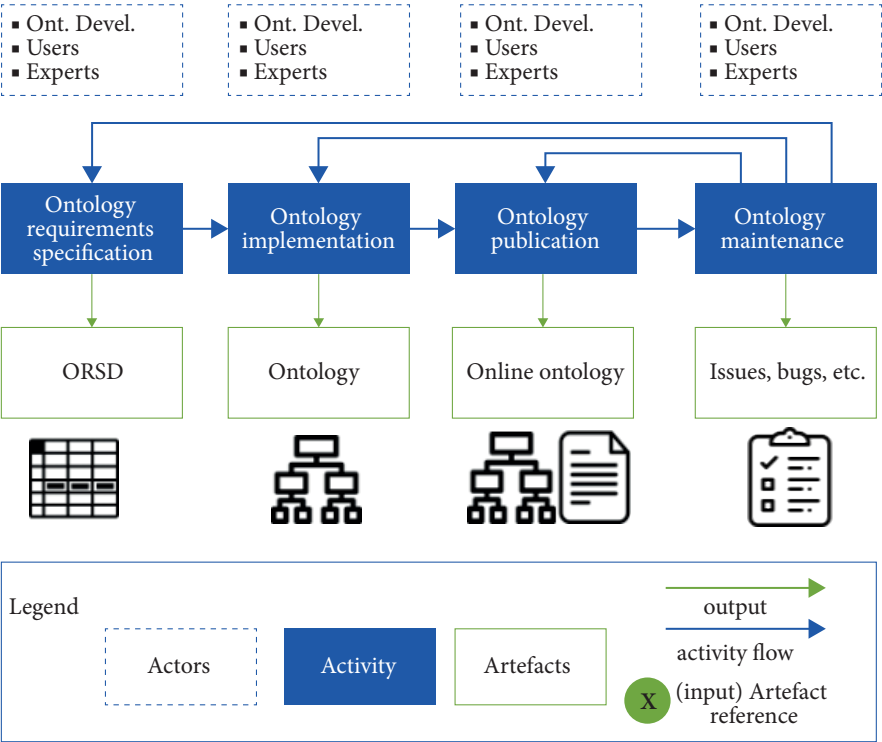
6. We refer the interested reader to the following authors for ontology development guidelines (Uschold and King 1995; Grüninger and Fox 1995; Gangemi, Pisanelli and Steve 1998), best practices (Bernaras, Laresgoiti and Corera 1996; Noy and McGuinness 2001) or self-styled methodologies (Fernández et al. 1997; Staab et al. 2001; Pinto, Tempich and Staab 2004; Suárez-Figueroa, Gómez-Pérez and Fernández-López 2012; Poveda-Villalón, Fernández Izquierdo and García Castro 2019).

7. First time mentioned in Grüninger and Fox (1994).

reproduced in the following (Suárez-Figueroa, Gómez-Pérez and Villazón-Terrazas 2009, 976):

- What is the job seeker’s education level?
- What is the job seeker’s work experience?
- What is the work experience required for the job offer?
- What is the economic activity of the employer?

The knowledge expressed in CQs gives ontology developers an initial idea of the main concepts and relations that are to be represented in the ontology. This is then complemented by a thorough analysis of available domain corpora, the identification of terms in documents, and the construction of a glossary of terms. Ontology engineers commonly rely on automatic term extraction tools for this step. The result of this extraction process is then validated by experts.



LOT Methodology <http://lot.linkeddata.es>
@Ontology Engineering Group

Figure 2. Activity workflow proposed by the Linked Open Terms (LOT) methodology

What this activity reveals is that the “natural point of departure” in the Ontology Engineering activity also consists of linguistic utterances expressed in the form of questions, statements, or descriptions in domain-specific literature. It bears many similarities to the initial phases of any terminography work, especially with the middle-out approach devised in Kerremans and Temmerman (2004) (see Section 3). The idea of an authoritative gathering of experts defining and naming the relevant concepts of a domain seems unlikely. Experts can influence the corpus to be used in the term extraction process, but the knowledge elicitation process can be defined as descriptive.

4.2 Knowledge-aware resources and knowledge patterns for ontology implementation

The next activity in the ontology development process is to implement the ontology using a formal language. At this stage of the process, most ontology development methodologies recommend searching for ontologies that cover a similar ground intending to guarantee interoperability. This activity is known as ‘ontology reuse’ and its purpose is to identify ontologies or ontology parts (ontology modules, ontology statements or ontology design patterns) that can be used again to minimize the time and effort involved in the ontology implementation activity.

With a view to automating the reuse of ontology statements and ontology design patterns (Gangemi 2005), there have been several efforts to identify lexico-syntactic structures in texts that can be directly mapped or translated into ontology structures (Cimiano et al. 2005; Cimiano and Wenderoth 2005, 2007; Aguado de Cea 2008; Montiel-Ponsoda and Aguado de Cea 2013). Some of this work has been inspired by previous research on knowledge-rich contexts (Meyer 2001; Marshman this volume) and term-relation extraction from text, and is framed in the broader line of research on ‘ontology learning’.⁸ This again shows the proximity of the methods involved in both terminologies and ontologies, and the impact that theoretical terminology studies have had on ontology work.

Even though ontology reuse is strongly encouraged in any ontology development process, recent studies have revealed that it is not such a consolidated practice among ontology engineers. The main reason for this is “the heterogeneity between the needed conceptualization and that of the available ontologies, as well as the deficiencies in some of such ontologies (concerning documentation or licensing)” (Fernández-López et al. 2019, 100492). It seems that the practice is usually reduced to some sort of ‘soft reuse’ in which only classes from certain popular

8. See Maedche and Staab (2001); Gómez-Pérez and Manzano-Macho (2004) and Cimiano (2006).

ontologies are reused. The authors mention the fact that more often than not there are contradictions in the intended meaning of terms between the imported ontology and the one being developed, and that the language of the developers and their cultural background also play a key role. More specifically, most ontologies are created in English and when used by developers in non-English speaking countries, adaptations are required to meet the needs of the target community.

What can be derived from this is that even if ontologies pursue the representation of a shared view of a knowledge area, the plethora of ontologies that are currently available shows that knowledge can be analyzed and structured from many different perspectives, and that there may be as many valid conceptualizations as communities of experts and uses of an ontology. This also reflects on language, when different terms denote very similar concepts or when there are various terms in one language to describe a phenomenon that has only one designation in another language.

4.3 Converting language resources into ontologies

Closely related to ontology reuse at this stage of the development process is the activity of ontology reengineering of non-ontological resources. Non-ontological resources are defined as “knowledge-aware resources whose semantics have not been yet formalized by means of an ontology” (García-Silva et al. 2008, 168). Examples are textual corpora, classification schemes, glossaries, thesauri, or lexicons. Whenever other knowledge resources are available for a domain, ontology developers may consider the option of reengineering them into ontologies (Villazón-Terrazas, Suárez-Figueroa and Gómez-Pérez 2010).

Villazón-Terrazas, Suárez-Figueroa and Gómez-Pérez (2010) provide an overview of existing methods for converting non-ontological resources into ontologies. The main idea is to map the resource schema into an ontology schema, and convert the resource content into ontology instances. To achieve this, the implicit semantics of the original resources has to be made explicit in the target ontology, which at a structural level means that most of the time, hyperonymy-hyponymy or meronymy relations are translated into subclass-of and part-of relations, respectively.

Although it is debatable whether such a direct correspondence should be established, it is generally accepted. The fact that this activity is part of the ontology development process indicates that on a practical level, a direct translation of lexical semantics into ontological structures is possible (although one could claim that some of the resources considered knowledge-aware resources, such as classification schemes or thesauri, lend themselves naturally to the same structuring principles as ontologies, contrary to what would happen with specialized dictionaries or lexica).

In this regard, L'Homme and Bernier-Colborne (2012) compared a domain ontology to a specialized dictionary to find differences and commonalities in the terms. They wished to discover if it is possible to exchange data between such resources. They concluded that there are major differences in the following: (i) focus of study (concepts vs. terms); (ii) the methodologies followed for their construction (the unambiguous definition of concepts vs. the specification of meaning, semantic relations and syntactic behavior); and (iii) the types of labels used as well as their assignment and purpose (artificially created labels for abstract concepts vs. lexical units directly extracted from text).

Even though ontologies and specialized dictionaries are different types of resources created for different purposes and users, and hence, exhibit undeniable differences, their similarities are worth emphasizing. The *de facto* (re-)use of non-ontological resources in ontology development proves that they may be more alike than previously thought. It would be worth quantifying the level of coincidence or overlap between specialized dictionaries and ontologies, to identify the type of lexical and semantic information that cannot be covered in ontologies. In this regard, Section 5 discusses models that aim at complementing ontologies with richer linguistic descriptions.

The abundant coincidences in the methods followed for terminology compilation and the first activities in the creation of ontologies are good evidence of the proximity of these two disciplines. This should encourage practitioners from both sides to mainstream some terminological and ontological practices in their respective methodologies.

5. Towards an enrichment of ontologies with terminological knowledge

One of the most important benefits of ontologies is that they enable interoperability between heterogeneous information systems. Interoperability is mainly achieved in two ways: (i) by agreeing on the knowledge to be represented; and (ii) by encoding that knowledge according to standard models and formats (ontology web languages). Thus, by complying with web standards, ontologies guarantee interoperability, reuse, and commitment to best practices.

In Section 2 it was mentioned that TMF and TBX are proposed ISO formats for the representation of computerized terminologies. Another ISO standard that attempts to mimic the underlying structure of electronic lexica is the Lexical Markup Framework (ISO 24613-1). As with any ontology standards, the main objective of these formats is to enable interoperability among language resources and facilitate their reuse by NLP applications.

The Ontology Engineering community was inspired by these standards when arguing that ontologies needed richer linguistic descriptions to be fully exploited by ontology-based NLP applications. The main objective was to take advantage of the wealth of knowledge that had been already conceptualized in ontologies in various domains, and to enrich it with a linguistic/terminological layer in one or several languages. At the same time, by representing linguistic information in Semantic Web standards, that information would become available for integration in semantic-based systems.

As detailed in Section 1, ontology languages only provided ‘labels’ and ‘comments’ to associate linguistic descriptions with ontology elements. These descriptors were unable to account for additional linguistic information, such as the morpho-syntactic behavior of ontology labels. Several models⁹ then emerged to interface language resources with ontologies as a crucial step for the use of ontologies in tasks that involved natural language (Huang et al. 2010). From those models, the ones that have gained adoption are *lemon*, the LExicon Model for Ontologies (McCrae et al. 2012), and its evolution, the Ontolex-lemon model. Ontolex-lemon is the result of the work carried out by the W3C Ontology Lexica Community group¹⁰ and was published as a W3C community report in 2016.¹¹

Ontolex-lemon is a general model that provides a framework for modeling linguistic information in ontologies, which is agnostic concerning the data categories or linguistic theory that can be used to instantiate the model. This model consists of a core (see Figure 3) and several (extensible) modules. The core classes provide a principled link between a `LexicalForm`, a word or term, and the concept it evokes for a certain ontology, which is represented by the `OntologyEntity` class. The `LexicalSense` “represents a set of disambiguated uses of an entry when used as referring to a certain concept”, i.e., the meaning of a word when interpreted with respect to an ontology (Cimiano et al. 2012, 18). This class allows users to represent pragmatic constraints and implications related to the usage of a given word when evoking the concept in question.

As for the rest of the modules that can be combined with the core module, they cover several linguistic features of lexical entries, namely, morphological decomposition, syntactic frames, and term variants and translations.

9. LingInfo (Buitelaar et al. 2006), LexOnto (Cimiano et al. 2007), LexInfo (Buitelaar et al. 2009), LIR (Montiel-Ponsoda et al. 2010).

10. Ontology Lexica Community Group (<https://www.w3.org/community/ontolex/>). Accessed 11 December 2021.

11. Lexicon Model for Ontologies: Community Report. 2016 (<https://www.w3.org/2016/05/ontolex/>) Accessed 11 December 2021.

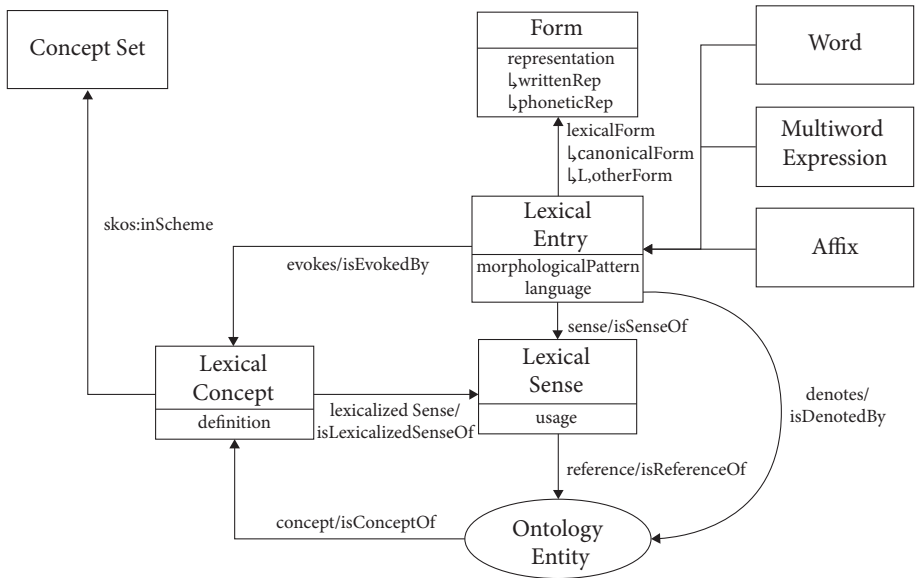


Figure 3. Ontolex core

Ontolex contains a module for the representation of term variants and translations, called the *vartrans* module and illustrated in Figure 4 (variation has been dealt with extensively in Terminology, see chapters in Part 7 of this volume). This module defines the class *lexico-semantic relation* (*vartrans:LexicoSemanticRelation*) and two subclasses (*vartrans:LexicalRelation* and *vartrans:SenseRelation*) to allow the representation of relations between terms of different nature. The *LexicalRelation* class encodes the relation between *lexical variants*, variants that are semantically coincident but formally different, mainly motivated by morphological and orthographical variation (localise vs. localize; ecological tourism vs. eco-tourism). The *SenseRelation* class encodes the relation between *term variants*, i.e., variants that are not only formally, but also semantically different.

According to Diki-Kidiri (2000, 29ff), in order to better understand this latter type of variant, it may be useful to make a distinction between concept and meaning, since these terminological variants refer to the same concept but represent “the multiple specific perceptions of the same object”, hence their classification under ‘sense variants’ (see Diki-Kidiri, this volume). Similarly, translation is understood as a ‘sense relation’ between words or terms that belong to different natural languages.

Term variants can be further classified into the following (although such a classification is not prescribed by the model): (a) diatopic (dialectal or geographical variants) (e.g., gasoline vs. petrol); (b) diaphasic (register-dependent variants) (e.g., headache and cephalalgia; H1N1 and the Mexican pandemic flu); (c) diachronic

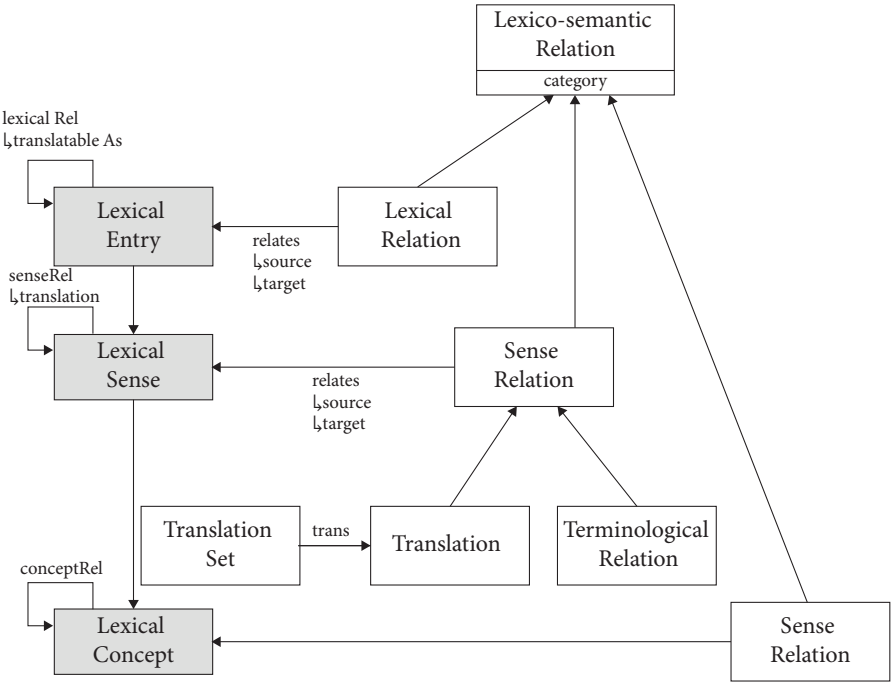


Figure 4. Ontolex vartrans module

(or chronological variants) (e.g., tuberculosis and phthisis); (d) diastratic (discursive or stylistic variants) (e.g., man vs. bloke); (e) dimensional variants: the terms point to the same concept but highlight a different property or dimension of the concept (e.g., biosanitary waste vs. hospital waste; Novel Coronavirus vs. Middle East Respiratory Syndrome Coronavirus; obsolete technology vs. dangerous technology) (Montiel-Ponsoda et al. 2013). By specifying the type of terminological relation, the reasons that cause that variation can also be captured by the model (diatopic, diaphasic, etc.).

One of the most interesting aspects of this model is that it is committed to W3C standards, which means that it is interoperable with ontologies and other language resources adopting the same format. In fact, Ontolex and its immediate predecessor (*lemon*) have become the *de facto* standards for publishing language resources in the Web of Data, and have been used by important resources in the Linguistic Linked Open Data cloud or LLOD (Cimiano et al. 2020), as, for instance, BabelNet (Navigli and Ponzetto 2012).

LLOD is an initiative to publish language resources (corpora, lexica, dictionaries, terminologies, thesauri, linguistic data categories, etc.) in open and reusable standards. According to the principles of this paradigm, any information item (e.g., term, definition, or part-of-speech) becomes a data resource, which means that it is

assigned a Universal Resource Identifier or URI and can be accessed and linked to another resource. The result of this linking is a huge graph of interconnected data resources, as envisioned by the fathers of the Semantic Web (Berners-Lee, Hendler and Lassila 2001). The LLOD cloud is part of the more general Linked Open Data or LOD cloud (Heath and Bizer 2011) (see Figure 5), in which data resources about Geography, Life Sciences, or Media are also published in web standards and are openly available. The fact that language resources follow the principles of this paradigm facilitates not only connectivity with other resources in the same formats, but also their integration in the most up-to-date applications that have embraced these standards.

As a proof of this concept, several terminologies have been converted to OntoLex-Lemon and interlinked in the TerminotecaRDF project (Bosque-Gil et al. 2016). In this project, the Terminesp terminological database (Wikilengua: Terminesp 2020) (containing more than 30,000 terms from technical standards with translations into Italian, English, German, French and Swedish), and several multilingual terminological databases in closely-related domains available from the Catalan Terminology Centre (Termcat 2020) (in Catalan, Spanish, English, and French) were connected in a common graph and represented as linked data on the Web.

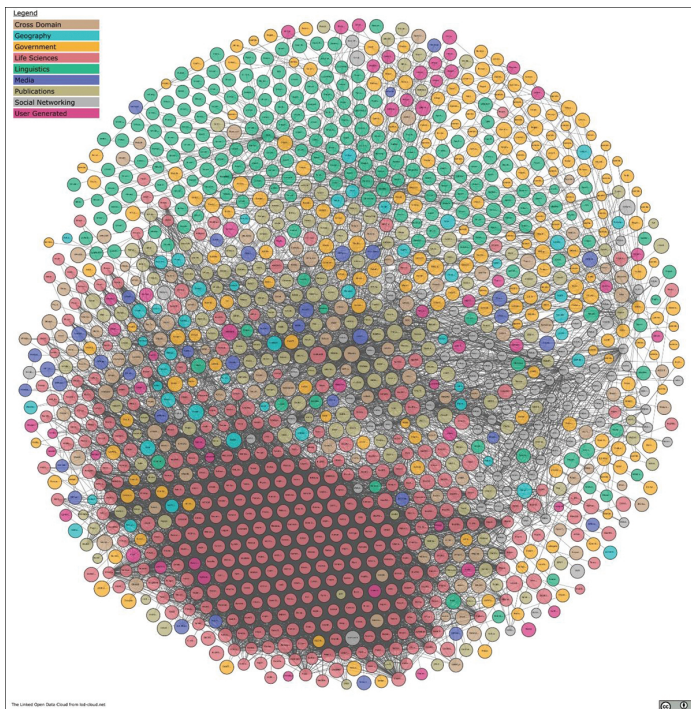


Figure 5. Linked open data cloud from lod-cloud.net

The main objective of this work was to demonstrate the added value of bringing together terminological data contained in distributed resources in different data formats (usually known as ‘data silos’). These data can now be jointly queried from a single access point. Moreover, these terminological data are currently ready to be integrated with ontologies of closely related domains, to further enrich their language layer and be used in ontology-based NLP applications.

6. Conclusions

This chapter has highlighted the common ground between terminologies and ontologies. After reflecting on the notion of ‘computational ontology’, we reviewed terminographic methods, tools and terminological resources organized conceptually to assess their proximity to ontologies. Additionally, such concept-based terminologies have been analyzed from the viewpoint of their potential conversion to Semantic Web formats and integration in NLP applications. Particular attention has been paid to the type of lexical and semantic information captured by those resources as well as how multilingualism has been handled.

Then, common practices in Ontology Engineering were described. We identified several tasks and activities in the ontology development process that are reminiscent of terminography, and which would greatly benefit from further terminological analysis. The final section highlights current trends and models in Ontology Engineering whose aim is to enrich ontologies with linguistic (terminological) descriptions. To avoid the limitations imposed by formal languages as well as the limited amount of linguistic information in ontologies, these models propose principled representation mechanisms to account for the lexical, syntactic, semantic, and pragmatic information relative to the knowledge captured in ontologies.

The two perspectives analyzed in this chapter, namely, the interest of ontological representations in Terminology and the inclusion of terminological analysis in Ontology Engineering, both support the idea of a renewed convergence between the terminological and ontological worlds. Ontologies are powerful artifacts to represent the conceptual structure underlying a specialized subject field. However, they fall short of capturing the lexical behavior of terms in text. In this regard, comprehensive flexible models to interface language resources with ontologies are still needed, in line with the ones described in this chapter.

In this sense, several challenges still remain open. One of them has to do with a wider adoption of these models and formats by other communities. Researchers and users outside the Ontology Engineering community demand simpler and more usable semantic technologies. Consequently, ongoing European projects, such as Prêt-à-LLOD (2020), are striving to create tools to support, amongst others, the

creation and conversion of terminological resources to the linked data format in way that is transparent to the user (see Terme-à-LLOD (Buono et al. 2020) and TermitUp (Martín-Chozas, Ahmadi and Montiel-Ponsoda 2020) as examples of such tools).

By sharing and linking their resources to the LLOD, terminologists, and linguists in general, would also benefit from the abundant resources already published in linked data formats. For instance, they could easily enlarge their terminological resources with translations in other languages from resources in the LLOD cloud. Specific query interfaces could be designed for language professionals as a single access point to information distributed in resources around the globe. They could also benefit from knowledge resources in the LOD modeling domain-specific data and integrate them in their NLP applications.

Ontology Engineering and Terminology are two disciplines deeply marked by interdisciplinarity, and each one has strongly contributed to the other. Relying on successful experiences in the past, we argue that paths need to be explored to identify further areas of common ground. One way to achieve this is through training. Specific terminology courses could be designed for ontology practitioners, and terminology extraction and management tools tuned and integrated in ontology editing tools. On the other hand, terminology practitioners should get special training on the use of semantic technologies to leverage LOD.

We envision a future in which terminological data will be shared and connected on the Web. This will undoubtedly lead to the creation of better NLP applications that will consume these data for the benefit of language professionals, hopefully, in the not-too-distant future.

Acknowledgements

The author would like to thank Guadalupe Aguado-de-Cea and the editors of this volume for their useful suggestions to significantly improve the chapter.

Funding

This work has received funding from the EU's Horizon 2020 Research and Innovation program through the contracts Lynx (grant agreement No. 780602) and Prêt-à-LLOD (grant agreement No. 825182), NexusLinguarum (CA18209), and from the Spanish Ministry of the Economy, Industry and Competitiveness through the Datos4.0 contract (TIN2016-78011-C4-4-R).

PART 3

Socioterminology and Cultural Terminology

Founding principles of Socioterminology

Valérie Delavigne and François Gaudin

Université Sorbonne Nouvelle / Université de Rouen

Socioterminology as a current of thought and academic research emerged in the 1990s. Research carried out in various fields has led to different theoretical approaches. Certain concepts were revised, and others were created. These new proposals made it possible to take into account the great variety of real language practices. This chapter reviews the development of Socioterminology, its actors, and its conceptual and methodological implications.

Keywords: history of Socioterminology, concepts of Socioterminology, variation, language practices, usages

1. Socioterminology: First steps

How Socioterminology emerged as a branch of thought and academic research has already been explained from its origins to the present (Gaudin 2005). Nevertheless, a little background knowledge is necessary to shed light on certain important characteristics.

Socioterminology came into existence because terminologists were dissatisfied with the limitations imposed by the General Theory of Terminology (GTT), whose primary aim was technical standardisation. These restrictions were particularly evident in the terminology methods used for language planning and were a subject of intense debate at the international conferences organised in Quebec by the *Office québécois de la langue française* and the *Ordre professionnel des traducteurs, interprètes et terminologues du Québec*.¹ Participating in these conferences during the 1970s and 1980s were well-known terminologists such as Pierre Auger, Jean-Claude Boulanger, Jean-Claude Corbeil, Louis Guilbert and Alain Rey *inter alia*. In fact, Eugen Wüster, the founder of the GTT, also presented a paper (Wüster 1976) at one of these conferences.

1. Order of Certified Translators, Terminologists and Interpreters of Quebec (<https://ottiaq.org>). Accessed 13 December 2021.

The term, *socioterminologie*, appeared sporadically at the beginning of the 1980s in Boulanger (1981) and Lerat (1984). Slodzian (1986) also used it with regard to Terminology in the USSR. However, it was only in Gambier (1987) that Socioterminology was finally given structure and content. Gambier, whose main focus was on interactional linguistics, had settled in Finland but had studied at the University of Rouen before Terminology studies were officially established there. At the time, the linguistic research centre at the University of Rouen was led by Jean-Baptiste Marcellesi who co-authored a guide to Sociolinguistics with his colleague, Bernard Gardin (Marcellesi and Gardin 1974), and forged the concept of 'glottopolitics' with another colleague, Louis Guespin (Guespin and Marcellesi 1986). This influential research team had initially worked under the supervision of Jean Dubois and Louis Guilbert, who was also an expert in Terminology (Guilbert 1973). Socioterminology was predominantly characterised by a dual approach that was sociolinguistic on a scientific level and materialistic on a philosophical one (see chapter by Humbley in this volume).

At the time, Louis Guespin seized the opportunity to work on a bilingual computing dictionary with a German partner. This is when it was first discovered that theories of Terminology could be more clearly envisaged through the prism of Sociolinguistics. Louis Guespin had a vast culture and a communicative sense of humour. The surprise and interest in the theories we were discovering, their idealism, epistemological naivety and incompatibility with the modern world, made for much hilarity. And in order not to mislead the reader, it must be said here that, before thinking and arguing, we laughed.

All of this information helps to contextualize our research. As Socioterminology disagreed with mainstream terminology theories, it had to explain its perspective. Assal (1992); Gaudin (1993) and Pierzo (1991) were working on their dissertations or PhD theses at that time, which meant they had to clearly formulate and justify their hypotheses. Other students followed, anchoring their reflections to specific topics, some of which also became theses. The majority of these students had studied Literature since the option of majoring in Linguistics was only possible in their third academic year. This evidently influenced their profile and differentiated them from translators.

This is how theoretical reflection took shape in the 1990s. Louis Guespin's premature death in 1993 marked a shift in this reflection, but his influence could still be perceived as new PhD students joined the centre and work continued. The different human encounters and research areas, which involved a wide range of fieldwork, were invaluable for comparing ideas, refining concepts, and gathering new proposals (Gaudin 2003).

Theoretical reflection also appeared in publications which provided the opportunity to build collaborative ties and expand different perspectives. Jean-Claude

Boulanger from Laval University wrote a preface in 1991 to a collection of papers (Gaudin and Assal 1991) in the special number of a journal. Monique Cormier (University of Montreal) immediately included this publication in the reading material for her course on Terminology. Two other journal issues were subsequently published in Belgium (Gambier and Gaudin 1993) and Quebec (Gaudin 1995). These were followed by the first book on the subject (Gaudin 1993). Conferences were held, such as the one in Quebec on the issues, theories, and practices of language planning (Dolbec and Troestler 1994) and a seminar in France on the implantation of official terms (Delavigne and Gaudin 1994). Simultaneously, directors and members of terminology organisations began to realise the usefulness of sociolinguistic approaches to Terminology. Socioterminology was making a name for itself, and its relevance stretched beyond the frontiers of the French-speaking community (Faulstich 1995; Messaoudi 2002).

The dawn of Socioterminology also coincided with that of other terminology theories described in this volume. Different voices began to propose alternative models, such as the Communicative Theory of Terminology (Cabr   2000) and the Sociocognitive Theory of Terminology (Temmerman 2000, this volume). These theoretical principles are either consistent with Socioterminology or are at least compatible with the socioterminological approach. Textual Terminology (Bourigault and Slodzian 1999; chapter by Condamines and Picton in this volume), which is close to socioterminological in terms of theoretical stances and methodologies, is regularly mentioned (see Condamines 2018b, 2018a; L'Homme 2020c). Other approaches, which did not catch on, include Psychoterminology (*psychoterminologie*) (Kim 2017, see *supra*). Other branches of Terminology are Cultural Terminology (Diki-Kidiri 2008, this volume), Ethnoterminology (Depecker 2013) and Pragmaterminology, which in many respects are similar to Socioterminology (Delavigne and Vecchi 2016; Vecchi 2016; Parizot 2014). Some contributions in this book bear witness to this.

2. A terminology closer to usage

The theoretical principles developed during this period were set out in easily accessible texts in French (Gaudin and Assal 1991; Gambier and Gaudin 1993; Gaudin 1995, 1993; Gambier 1987, 1991b). These principles were mainly a reaction to the fact that the GTT did not take into consideration the fundamental role of interaction, which was in evident contradiction to mainstream Sociolinguistics. As an integral part of language practice, terminology also develops the language that produces it. Indeed, the consequences of this process have yet to be fully understood and are worth examining in greater depth.

Initial questions and initial questionings targeted the biunivocity of terms, the assumption of their monosemy within a particular domain, and a conceptual approach coupled with an onomasiological analysis (Gambier 1987; Gaudin 1989).

This debate first took place within the larger framework of glottopolitics (Marabini San Martín 2019; Morgenthaler García and Amorós-Negre 2020). This allowed researchers to simultaneously examine both the genesis of spontaneous norms as reflected in term usage and the influence of institutionalised standards. The principles of glottopolitics made it possible to pinpoint specific features pertaining to institutional standardisation of the various factors that influence speakers. It is here that phenomena such as self-hatred and linguistic insecurity, which have been studied in detail in multilingual contexts, were incorporated into terminology studies. Similarly, this interactional perspective ensures that lexical forms are always analysed in the context of the communicative acts in which they occur, even if some harmonisation may later be required to achieve mutual understanding.

Various studies were conducted by researchers whose work initially focused on standardisation but who quickly realised that the realities of communication required a change of method. A case in point is a comparative analysis of Vietnamese building terminology for French LSP classes carried out by Nguyen (2013). After initially applying Wüsterian principles, Nguyen decided to change her methodology when she discovered a significant disparity between the terms used in academic writing and those used at building sites. The analysis of the designations showed that 155 of 463 standardised terms had at least one alternative to the term actually used in practice. More generally, variants were found in all the sectors whose terminology was analysed, even when it could reasonably have been assumed that rules for naming were easy to apply. Another study showed that airline employees replace the strictly standardised names of airports with affectionate nicknames when flying between continents (Lopez 2013). The need for speakers to make language their own and to mark it with their identity seems to be a much more powerful force than otherwise thought.

Peraldi (2011) showed that despite the existence of well-established standardisation bodies in the field of Chemistry, the vocabulary in use is far less regulated than one might think. Some of the chemical components that have an official designation in the Chemical Abstracts Service (CAS) databank have been named differently by the International Union of Pure and Applied Chemistry (IUPAC), “ce qui constitue, une difficulté supplémentaire en termes d’appréhension du domaine et des concepts clés”² Peraldi (2011, 98). Moreover, the author points out that aside from these international standardised lists, each of these terms also has around

2. This complicates matters in terms of understanding a domain and key concepts (our translation).

ten terminological variants. This analysis does not address interdisciplinary fields such as Bioinformatics where identical subjects are compared from different perspectives, which means dealing with problems related to diverging classifications (Gaudin and Bouveret 1997).

When major security issues are at stake, terminology is expected to be more stable. Yet the analysis of the way nuclear production terminology circulates in different public sectors shows that terms vary in response to discursive strategies that are designed to promote or understate, depending on the communication issues at stake (Delavigne 2001). Such terminology has the particularity of being at the core of social issues that are directly related to progress and danger. This has impacted and still heavily influences the frequency and use of terms in this context (Delavigne 2006). It highlights the fact that, aside from the terms themselves, it is crucial to examine how they are used in discourse, which is why Socioterminology is grounded in discourse analysis.

In international diplomacy, term variation comes as less of a surprise, although the study of debates within the United Nations concerning human rights reveals various political, philosophical, and religious issues. However, the muted language used in such organisations is unable to gloss over the extent to which parties disagree. With regard to debates on Article 18 of the Universal Declaration of Human Rights concerning the right to freedom of religion, “pour une partie des diplomates, cette liberté incluait la liberté de changer de religion, alors que pour les représentants des États islamiques la Shari’a interdit l’abandon de la religion musulmane.” (Leroy 2005: 223).³ This variation covers strong ideological disparities.

In healthcare, the description of variation is crucial for optimising communication between the medical staff and patients. The diabetes epidemic in Kuwait, for example, ran into difficulties stemming from the use of competing terms that are often ill-adapted to the general public. This terminological breakdown is linked to the fact that doctors are trained in different languages, though mainly in English. When caring for their fellow citizens, they do not have access to the terminology they need for their patients to understand them. This results in terminological fragmentation which would require careful adaptations to facilitate both spoken and written doctor-patient communication (Alrashidi 2016).

Issues in Oncology are slightly different as the choice of words in such an anxiety-inducing context carries particular weight. Obscure vocabulary and the fast-moving developments in treatments and channels for disseminating information are all factors that deserve close attention if patients are to gain important

3. For some diplomats, this included the right to change religion, whereas for representatives of Muslim countries, abandoning Islam was prohibited by Sharia law (our translation).

knowledge and information about their condition and become active participants in their treatment (Delavigne 2012a, see Section 5).

In a research study on Igbo vocabulary in obstetrics and gynaecology, Ngwaba (2016) examines gaps in vocabulary and then proposes guidelines for the creation of new terms. These proposals are made by taking into account the historical dimensions of the terms used. For the author, the issue involves “not only creating terms, but expanding the Igbo language, reviving forgotten words and asserting Igbo’s capacity to name concepts in the field [of obstetrics and gynaecology]” (Ngwaba 2016, 155). To achieve this, she conducted a large number of structured interviews to draw up an inventory of terms. Such questions are vital for public health. The new information is added to existing mental images, and the terms used must convey meaning – a method that was successfully used a quarter of a century ago for sight-impairment pathologies in the Sahel (Slodzian 1994).

The socioterminological approach often comes into play when greater emphasis is placed on describing variation for purposes of language policy. Kim (2017) argues that socioterminological factors of implantation coincide with these criteria and are not consistent with terminological, psychoterminological, and extraterminological factors. From her point of view, these socioterminological factors “relèvent de la relation établie entre le terme et la société. Ils comprennent, plus particulièrement, l’aspect institutionnel de la société, à savoir toutes les activités effectuées au niveau de l’aménagement terminologique”⁴ (Kim 2017, 135). Further on, she points out that “Ces facteurs ne concernent pas la structure intérieure d’un terme, mais plutôt le lien entre un terme et l’environnement, c’est-à-dire la circulation des termes dans la société”⁵ (Kim 2017, 142).

In Terminology theory, examples of term variation are plentiful. A case in point is the coexistence of the terms, *concept* and *notion*, which are reflected in the title of Rey’s (1979) short work, *La terminologie: noms et notions* (1979), in which they are used interchangeably. *Notion* became *concept*, which sometimes clashed with its epistemological meaning of theoretical concept, i.e., the onset of scientific thought. Polysemy in terminology is a reality that is frequently acknowledged. Even the word, *term*, is subject to variation as it oscillates between *term*, *specialised term*, *technical term*, *terminologism*, *terminological unit*, and *specialised lexical unit*. Evidently, variation is a constituent part of all language practices, however terminological they may be.

4. [socioterminological factors] are founded on the established relationship between term and society. More specifically, they include the institutional aspect in society, namely the activities undertaken for terminology planning (our translation).

5. These factors do not concern the way a term is formed but rather the relationship between the term and the environment i.e. the circulation of terms in society (our translation).

Sociolinguistics reminds us that there are many types of variation. The most obvious one is diachronic variation which reflects the state of a discipline's conceptualisation (i.e., *notion* and *concept*). Guespin (1991) took pleasure in alluding to *germs*, *microbes*, and *micro-organisms*. It is inconceivable to study vocabulary without considering how it developed, a point highlighted in Dury and Picton (2009). However, the other types of variation (diatopic, diastratic, diaphasic, diamesic, etc.) are all relevant to Terminology when the analysis of language practices is the purpose (Delavigne 2017; see chapters in Part 7 of this volume).

3. Concepts deployed

To provide answers to the questions raised by the GTT, Socioterminology borrowed some notions from other disciplines and forged new ones. One of the first questions that arose concerned the domain. To make room for interdisciplinary and horizontal collaboration, Guespin's (1991) *interface discussion* is better called *sphere of activity*, which highlights the continuum between science, technology and production. Activities are often carried out in different specialised sectors. Their interaction leads to mutual development and vocabulary follows. *Sphere of activity*, which Guespin borrowed from Bakhtine, neutralizes differences between approaches, such as science, technology, production, and marketing, which are in constant interaction and communication. From this perspective, terms are not regarded as mere labels but rather as linguistic units that circulate and have different nuances, depending on the utterances and the type of texts in which they appear.

As with language contacts, contacts between different branches of knowledge can require adjustments. When examining verbal interactions in language practices, certain negotiations that people undertake in an effort to adapt to each other become apparent. These negotiations sometimes concern the choice of words, and are known as *naming negotiation* since the speakers must agree on the linguistic signs. They can also involve discrepancies between concepts, known as *cognitive negotiation*, when disagreement arises because of differences in culture, language, or discipline.

These adjustments are particularly prevalent in science popularisation. The sociolinguistic perspective is rooted in the study of the way terms circulate in society. Its mission is to foster a deeper understanding of the mechanisms that cause terms to appear in different spheres of activity, as well as their usage and dissemination (Gaudin 1995). Texts that popularise science are valuable resources for fieldwork in this area. These discourses involve the use of specialised linguistic forms while avoiding the comprehension difficulties that they often cause. When such difficulties arise, it is important to determine whether the problem lies with the signifier or the signified.

The greater the number of unknown terms in a text, the greater the amount of new information in the text, and the less likely the text is to retain the reader's attention. Conveying information to a broad audience implies mastering semiotic tools and above all terminology. Exceptional circumstances, such as the AIDS or Covid-19 pandemics, require a great deal of communication about highly technical notions so that entire populations are able to understand the issues and can thus adopt a rational attitude. Nuclear accidents, genetic manipulation, discoveries in astronomy, viruses, extreme climatic conditions and political, economic, and social events generate a profusion of terms in the media (see Moirand 2014). The usages and 'designation paradigms' (Mortureux 1993) of these terms vary, depending on the situations and interactions. Their analysis requires switching to flexible descriptive models in order to better understand how terms reflect real-world objects in different speech communities.

Other resources can be used to examine how terminologies circulate, particularly those within the framework of the language planning tools and methodologies available. Like all public policies, political initiatives need to be assessed. Although research groups in France were invited to carry out pioneering studies on term implantation (Depecker 1997), more research is needed in this area (Quirion 2011; Renwick 2018). Research on the success or failure of term implantation coincides with more general questions regarding the development of usages and the sociology of lexical innovation (Martin 1994).

The circulation of terms can be studied by means of language networks and concepts, such as a 'linguistic bridge' (Corbeil 1980), whether for internal or external communication. This makes it possible to detect and identify sensitive interchanges in an organisation's communication which language experts can analyse and then try to find solutions for. Software applications for tracking neologisms provide support for automatic text analysis but little is known about how they spread orally, and which generally takes place first.

Questioning the concept of domain signifies examining the role of experts, who have the ultimate responsibility of determining which term falls within a specialised sector (Otman 1993). The role of experts is crucial in order to guarantee the accuracy of the resources proposed. However, aside from the fact that the notion of the expert is questionable, in Terminology as in democracies, the role that experts play in the way linguistic signs circulate should be clarified. This has been facilitated by concepts borrowed from the Philosophy of Language.

A look at the channels through which terms circulate shows that speakers entrusted with the responsibility of defining their 'legitimate' meaning occupy a privileged position. Even though they are called *experts*, that does little to further our knowledge. What really matters is the way in which all speakers are able to refer to a source in an accurate and consensual manner. If we are uncertain about the

meaning of a word, its definition, or its classification, we consult either an expert or an expert source.

Reference, which is central to Terminology and Semantics, can therefore be broached from a sociolinguistic perspective. When communicating, people produce terms whose effectiveness (i.e., the way in which they facilitate discussion and democratically share knowledge), presupposes that the speech community entrusts a very restricted number of sub-groups with the responsibility of determining their meaning. Moreover, these sub-groups must assume the responsibility conferred on them even in a crisis.

For example, with the appearance of exoplanets, no one was able to define the term, *planet*. This highlighted both the fragility and necessity of scientific consensus. Astrophysicists were obliged to make up their minds in order to fulfil their mission as guarantors of accurate references and had to vote on whether Pluto should be classified as a planet. Since August 2006, Pluto is no longer a planet in the solar system but just a dwarf planet (Gaudin and Nicolae 2016). Crises in international bodies, such as the one experienced by the International Astronomical Union, highlight the importance of the decision-making role that experts assume on behalf of society. Philosophers call this type of intellectual trust ‘deference’, this ‘referential delegation’ being the result of what Putnam (1975, 146) calls the “division of linguistic labour”.

In fact, when communicating, we agree on the meaning of terms generally by determining the reference. In other words, we cooperatively undertake a *referencing* activity when speaking to each other. The dimension in which everyday names are attached to realities that are put into words is a social one. The construction of the world through language, as we know and experience it, is therefore achieved through cooperative effort (Alrashidi, Delavigne and Gaudin 2018).

This cooperation is the basis for the security that we feel when communicating. More specifically, we need a ‘guarantee’, namely, the endorsement of normative bodies (Rastier 2011). However, because of differences in languages or disciplines, terms sometimes have to compete with each other even before their semantic differences have been clarified. This can produce both linguistic and cognitive insecurity (Gaudin 1994) because doubts about meaning create uncertainty regarding conceptual differences. This provides still another argument in favour of harmonising vocabularies, which are sometimes regulated spontaneously – a phenomenon that sociolinguists call *normaison*, in other words, a spontaneous norm (Guespin 1993).

The consideration of all discourses in which terms are used from a sociolinguistic perspective makes it possible to identify a social evolution of expertise. In some areas, this can lead to a redistribution of roles. The transformation of the ways in which information circulates in healthcare, for example, has blurred the boundaries between professional and patient knowledge. We now see discursive

digital communities developing a paramedical culture and terminological literacy that once more raises questions about expertise (Delavigne 2019b, 2020b).

This issue is in fact intimately related to the conception of the terminological unit. Traditional Terminology has brought the semiotics of terms into opposition with the semiotics of words. The question is whether terms, as specialized knowledge units that enable understanding and action, should not primarily be regarded as constructions whose specificities are linked to the status that a given community of discourse attributes to them.

4. Methods and analytical tools

Examining terminology from a sociolinguistic perspective requires a methodology for carrying out the descriptive program that Socioterminology has set out to achieve. Core issues for the description of usage include the following: the question of corpora as an observatory of language, methods for collecting data, and the way in which term usages are to be reinserted into the contexts of their production, circulation, and reception. Traditional terminology work begins with an onomasiological approach. Although a conceptual map of a specialized domain can sometimes be useful for organising the terminological workflow (Dubuc 1992; Pecman 2018), does it illustrate language practices? In any case, criteria must still be established for the selection and organization of corpora.

For written resources, Baudet's (1991) editology is an epistemological system that can be used to replace abstract notions, domains, and disciplines with facts, such as the scientific or editorial committee that validated a text, the manner in which it was published, the means of distribution, and the author's expertise. Both texts and terms are disseminated through channels that are objectively understood. 'Editological records' describing the collected texts can therefore be created with a view to placing the terms back into the contexts in which they appeared (Delavigne 2001, 2008; Delavigne et al. 2007).

Organising corpora into different types implies an in-depth knowledge of fieldwork, a notion which, as Boutet points out, concerns all linguistic practices that address "methodically and theoretically, the social situations in which the language material under analysis was produced" (1994, 2–3). In general terms, sociolinguists consider that fieldwork is central to both the discipline and research, regardless of the methods deployed for gathering data. By proceeding descriptively, Socioterminology borrows the methodologies of Sociolinguistics (in this sense, *socio-* stands for sociolinguistic). Fieldwork has led researchers to question these sociolinguistic methodologies and to adapt them in order to address real-life term usage and what it represents for the speakers, depending on the social situations

of the utterances (Gaudin and Delavigne 1995; Delavigne 1996; Delavigne and Gaudin 1996).

In the Digital Age, access to domain-specific texts seems to be obvious. The explosion of the Internet has profoundly altered ways of gathering language data and their analysis (Tanguy 2013). However, gaining access to documents circulating within a company, for example, is not always easy for reasons of confidentiality. For some spheres of activity, methodologies for collecting texts have not as yet been developed (Delavigne and Vecchi 2021).

However, written sources are only one part of the picture. Socioterminology has highlighted the importance of spoken language, which defines the majority of interactions and is the medium through which terms circulate in reality. Aside from a few specific semiotic units such as symbols and other mathematical formulas, terms mainly circulate in oral communication. If the objective is to build terminologies based on actual usage, it is necessary to examine the way terms really work. This means focusing on communication that includes all the forms of a term that may differ when spoken because of the multiple registers of language. Apocopes, for example, are frequently used in speech but often compete with longer forms when written. *Anapath*, which can be considered as a term in its own right, competes with the longer form, *anatomopathologic* in certain texts, such as medical reports (Delavigne 2017).

Progress in Corpus Linguistics has greatly facilitated the way that sociolinguistic corpora are compiled (Condamines 2005; Williams 2005; Mayaffre 2005; Rastier 2001). Today's input from native digital corpora, such as blogs, forums, and other social media, provides valuable sociolinguistic data. Despite the specifics of their prose, they are now important sources for examining terminologies in use and can be the subject of a "netnography" (Kozinets 2015). They represent a range of exploratory areas whose *technodiscursive* dimension, a word coined by Paveau (2015), reconfigures language usages. As areas of *normaison*, they have become a terminological resource in their own right, where variations and meaning adjustments can be located (Delavigne 2013, 2019a). The interactive exchanges deployed there provide a very rich terminological repository that has yet to be fully investigated.

Whatever the type of corpora that is used, the analysis of terms and their circulation is inseparable from the requirements and issues related to them, which differ depending on the fieldwork. They may involve language planning, terminology creation, terminographic description, resource development, and recommendations for specialists, among others. Socioterminology emerged from a critical stance *vis-à-vis* Terminology whose focus on standardisation reflected an idealised dream of unambiguous language in which term usage was carved in stone and not susceptible to variation. This was detrimental to the description of how language is used in real life, which involves the multiplicity of its interactions. In a nutshell,

the General Theory of Terminology (GTT) “dicte, puis se lave les mains”⁶ (Guespin 1993, 10). In the light of fieldwork, this critical stance was corroborated by the need to respond to social demand, because aside from descriptive analysis, the core purpose of Socioterminology is to be of use to society.

From the outset, terminologists have been faced with putting theory into practice, which is not the case for most theoretical linguists. Socioterminology considers that description takes precedence over prescription, and that its application is equally as important. Socioterminology has a dual vocation, “d’une part décrire, d’autre part informer”⁷ (Gaudin 1991, 178) from a *glottonomic* perspective. *Glottonomy*, a term coined in Guespin (1985, 26) is defined as follows:

La glottonomie propose, à la suite de l’analyse d’une situation langagière particulière, les modalités d’une intervention sur les pratiques langagières propres à cette situation.⁸

This vocation involves using description to understand, at a particular point in time, the parameters of a linguistic situation from which ways to improve communication can be inferred. The approach is supplemented with an advisory attitude, which can at times be difficult to sustain and which raises a certain number of questions about social responsibility (Delavigne 2015).

The application of a theory determines the value of the theory itself, i.e., its capacity to summarise facts and resolve problems. Meeting these requirements demands knowledge of the related issues. Variations in the use of terms are not random. They reflect circumstances, the role of the speakers, and their position in regard to others in the communicative context (client *vs.* vendor, doctor *vs.* patient, patient *vs.* user, etc.). The choice of terms is determined by diverging factors that can be highlighted through discourse analysis. Socioterminology particularly draws on the *terme pivot* [pivot-term] method. As a gateway into a text, the term becomes a pivot point, which is reflected by its lexical and predicative links with the other units in a given corpus. These co-occurrences are then connected to the speakers, the context in which they appear, and the circumstances of the discourse. The end result sheds light on the discourse under analysis.

6. [The GTT] dictates, then washes its hands of the problems (our translation).

7. To describe and to inform (our translation).

8. Once a particular language situation has been analysed, *glottonomy* offers methods for intervening in the language practices pertaining to this situation (our translation).

5. Socioterminology in practice

Socioterminology is situated at the heart of speech communities. Its aim is to provide descriptions as well as terminological resources such as glossaries, dictionaries, and databases. Its guiding principle is to optimally observe term usage with a view to developing useful and flexible solutions that reflect language practices. This focus on usage means that Socioterminology must include the whole range of term variations that circulate in and between discourse communities.

As an illustration, we describe the compilation of a dictionary for cancer patients (Delavigne 2012a, 2012b; INCa 2013). The dictionary was part of a health policy that encouraged the active participation of patients in decisions about their therapy. This evidently involved the appropriation of the terms. It also responded to the desire of patients and their families to obtain reliable, validated, comprehensible and up-to-date information in tune with their needs. The patients and their families could consult the dictionary and look up the terms that they were most likely to encounter.

On a methodological level, this required a previous in-depth analysis of current discourses and knowledge demands about what makes sense, based on specific needs. Although it is often assumed that medical matters are confined to only one domain, nothing could be further from the truth. Of course, a medical community is composed of professionals such as doctors, nurses, surgeons, chemotherapists, radiotherapists, etc. However, cancer involves other professionals who are also in contact with patients, such as social workers, hospital beauticians, charity workers, and more. Cancer patients meet a mixed community with whom they not only discuss their illness but also their daily lives, psychological difficulties, social problems, and work. Relevant terminology is thus not only strictly medical but also belongs to other domains.

For example, patients and their families also need to use and understand legal terms or those relating to social work, such as *tutelage* or *legal guardianship*, psychological terms, such as *denial* or *addiction*, as well as those related to the organisation of cancer treatment, such as *personalised cancer care*. Domain boundaries thus become fuzzy and unstable. Cancer-related terms often appear outside strictly specialised spheres. Patients do not only talk to medical teams, but also with their families. They read the media, watch television, consult Internet and are active on social media. A constellation of interwoven discourses emerges in which multiple voices communicate, discuss, and interact with each other.

The corpora thus had to be expanded in order to include these different areas in which terms circulate and proliferate, and where multiple voices and intertextuality mingle. Here, the notion of expertise comes up for review as it is no longer linked to validation by a certain institution. The analysis of corpora from cancer

patient forums highlights a community of patients who have mastered specialised terminology and possess what is known as “paramedical culture” (Delavigne 2020b, 2013). This is demonstrated in patient utterances such as the following:

Le SYD985 est un immuno conjugué liant un anticorps anti HER2 (de type Herceptin) à une à une toxine cytotoxique.⁹

This example uses updated acronyms, abbreviations and other terminological units in association with metalinguistic adjustments. Traces of naming negotiations or *routines discursives* [discourse routines] (Delavigne 2020a), such as *talk about* in reference to the term, *remission*, are evident in the following extract:

Quant à la notion de rémission, mon oncologue parle davantage de réponse complète, c’est à dire qu’il n’y pas plus (à l’analyse des examens cités) de lésions suspectes, tout au moins on en voit plus (c’est comme ça que mon oncologue l’exprime..)

En effet, elle dit que pour les cancers métastasés, on ne parle plus de rémission, la maladie est chronique même si on ne repère plus de lésions suspecte.¹⁰

Such abundantly intertextual excerpts provide evidence of expertise that has been built up through regular interaction with the medical world. New expert figures emerge in the form of people who share their experience (and whose capacity to popularise specialised language is assessed as superior to that of health professionals). This is validated by other patients:

Le chirurgien m’a bien expliqué tout ça et l’anesthésiste aussi mais ton témoignage est nettement plus clair.¹¹

Experts are no longer necessarily from the world of academia. They are appointed by their peers, and their experience has established a new form of validation (Delavigne 2020b).

The concept of ‘genre’ (see Pecman and Kübler, this volume) also needs to be addressed, particularly in regard to hybrid discourses. This issue also arises when analysing new types of discursive text, such as new Internet genres (if they can be

9. SYD985 is an immunoconjugate binding an anti HER2 antibody (type Herceptin) to a cytotoxic toxin. Posted on 11/06/2015 by bernadou1 (our translation).

10. As for the notion of remission, my oncologist talks more about complete response, in other words there are no more (in the results of the examinations mentioned) suspicious lesions, or at least none have been detected (that is how my oncologist puts it). In fact, for metastasized cancers, she says that we no longer talk about remission, the condition is chronic even though there is no longer any trace of suspicious lesions. Posted on 31/08/2015 by mimia (our translation).

11. The surgeon explained all that to me, and the anaesthetist as well, but your explanation is far clearer. Posted on 01/02/2013 by Ourse polaire (our translation).

called *genres*) such as blogs or forums. When analysing Internet discourses, new categories must be established (Mourlhon-Dallies, Rakotonoelina and Reboul-Touré 2004; Panckhurst 2007; Anis 2009; Liénard and Zlitni 2015; Paveau 2014).

These corpora, which are a mixture of polyphony and intertextuality, are also moving discursive spaces. For example, news about the Covid-19 pandemic, which has been produced in an already anxiety-inducing context, has caused an influx in Internet forums of terminological units from related domains. In a very short time, media has opened up a vast workshop of neologisms, and social media has caught on. Although data on research methods, such as *clinical trials phase I, II or III* had already been identified in these corpora, terms, such as *social distancing*, or *physical distancing*, and many more, have also made their appearance.

In this type of scenario as well as many others, corpus analysis makes it possible to closely observe both the practices and needs of language users. Term variation and semantic negotiation are located there, as well as epilanguage, when, for example, a term seems complicated or scary (Delavigne 2020a). As a consequence, these variations should be analysed, described, and modelled.

However, one wonders what becomes of terms in these language markets. In such polyphonic discourses suffused with intertextuality, the identification of terms based on their membership in one domain or another (which we are at great pains to determine) is no longer viable. Even the ‘sphere of activity’, which was initially advanced by Socioterminology hardly seems relevant any more. Domain boundaries have become irrevocably blurred. The terminological landscape is no longer modelled on domains, however numerous they may be, but rather on communication issues. Evidently, another discourse economy, with specific lexical choices and enunciation stances, has come into play. This influences those specialized knowledge units regarded as terms, whose semiotic status varies, depending on the discourse (Petit 2001). This obviously affects the very definition of the terminological entity. As Rastier (1995) observed, “le régime terminologique des textes varie selon les genres et les discours.”¹² A term can be identified as such only through its semantic value in a given situation. The previously mentioned question of whether anapath should be considered a term is revealing. It is the same with *integration* and *inclusion* which seem to be synonymous but in terms of disability take on different meaning depending on the discourse under consideration (Sivilotti 2019).

Socioterminological logic sets aside the axiom that concepts are primordial and primarily implements a semasiological and interactional approach. Terms are only terms when they are part of a discourse. The circumstances in which they are produced, their circulation, reception, and interpretation are the factors that confer ‘termhood’ on a lexical unit or multi-word expression.

12. Terminological regimes vary depending on the genre and the discourse (our translation).

6. Conclusion: Dissemination of Socioterminology and today's challenges

The introduction of a new term – like *Socioterminology* – always raises the question of whether it is really useful. Indeed, it is not uncommon in scientific discourse, the use of a new term may be of questionable relevance. It may have arisen because of the need to signal or capture a new concept or to more narrowly delimit the meaning of another. The new term can also serve as a banner but is always more effective as signifier rather than a signified because of its insufficient notional structure. Lastly, the usefulness of a new conceptual differentiation does not necessarily last.

As for the field of Terminology, theoretical stances have evolved considerably over the last thirty years, to the extent that the *socio-* in *Socioterminology* can almost be considered redundant. Today, many more linguists are addressing terminological issues. Moreover, the convergence of Terminology with Corpus Linguistics has given rise to approaches that are similar to those used in discourse analysis (Condamines and Picton, this volume). The consequences of pioneering empirical studies have coincided with the developments associated with theoretical studies, as reflected in the fact that there is now an entry for *socioterminology* in Dubois et al. (1994) and Charaudeau and Maingueneau (2002). ISO has even created a Socioterminology standard (ISO 2007),¹³ which reflects that it is an acknowledged area of research.

After a long period during which Terminology sought to distinguish itself from Linguistics, the gap gradually seems to be closing (Mejri, Petiot and Petit 2005). Some sociolinguistic methods have been adopted such as the following: (i) the use of corpora to build terminologies that reflect actual term usage; (ii) the analysis of term variation; (iii) the inclusion of the social uses of terms; and (iv) the study of diachronic and diastatic variation (Dury and Picton 2009). Today, these sociolinguistic dimensions have been generally adopted by common consent, which demonstrates the need to renovate the 'classical' Terminology of the late 20th century. Such reflections do not all originate from Socioterminology, but they do tend to share the majority of the criticisms that were the driving force behind its emergence.

The theoretical and methodological stances of Socioterminology have spread throughout the research community (Bordet 2013). Traces of its dissemination are found in a wide range of important publications (Cabr   1998b; L'Homme 2020c; Humbley, Budin, and Laur  n 2018). Nevertheless, the statement that the sociolinguistic dimension has become commonplace in Terminology and that it influences all aspects of its theory and practice is somewhat of an exaggeration. There is still much research that lacks a strong sociolinguistic perspective.

13. This standard is currently under revision.

Although the significance of Socioterminology has been widely acknowledged, it motivates few studies outside academic settings. Substantial longitudinal research including oral communication and the comparison of language practices in different domains or spheres of activity are the exception, aside from the previously mentioned PhD theses and Collonval (2021) on hospital communication. Internationally, the echo of standardisation-inspired Terminology based on theories originating from Eastern Europe still hold great sway. Although some progress has been made, the sociolinguistic dimension still has a long way to go before it is fully integrated into contemporary approaches.

Regarding the dissemination of the term, *Socioterminology*, a search on Google Scholar indicates its occurrence in 1,500 papers, 600 of which are French. In these publications, 10% present the variant *Socio-terminology*.¹⁴ A closer look shows us that it is difficult to identify researchers who explicitly claim to be socioterminologists. It is more commonplace for research to mention the contribution made by Socioterminology, whether theoretical or methodological.¹⁵

As Mortureux (2000) pointed out, it is difficult to use terminology by omitting the sociolinguistic dimensions of specialised discourse. In spite of significant progress made within the framework of Terminology, there is still a long way to go before this dimension can be considered to be fully integrated into contemporary approaches.

This is the main challenge in Socioterminology today. Observational evidence is no longer focused on individual terms, but rather on corpora. In other words, studies now mainly consider terms *in vivo*. Corpora, whether integrated or accessible, have irrevocably transformed terminology work, which is a very positive outcome. The issue of terminological corpora thus takes on another dimension. Within this broader framework, relevant questions include the compilation of a field terminology from which potential sociolinguistic issues can emerge. This field terminology should also facilitate the identification and usage description of oral and written term variants. It should also provide a rich context in which their projection and circulation can be modelled.

As we know, terms are not only updated in peer-to-peer discourse. A term, or word is only 'right' when it is appropriate in a given communication situation, even an asymmetric one. It is thus these uses that should be identified and described by

14. This search was run on September 25, 2020.

15. This is the case with a large number of researchers like Anne Parizot into the terminologies circulating within the multinational Michelin (Parizot 2014) and Laurent Gautier at Bourgogne University; Anne Condamines at Toulouse University; Maria Francesca Bonadonna and Rachel Raus (2013) in Italy; Maria Teresa Cabré in Barcelona; Aurélie Picton at Geneva University; or Leila Messaoudi in Morocco, for instance.

placing them in their ecological environment. This imperative means taking into account the diversity of discursive genres produced by communities, both internally and externally.

However, it is necessary to decide what exactly should be collected and retained, and how discourses can be mapped and text genres typified. The question is whether it is useful to process vast amounts of data in which it may be difficult to circumscribe specific phenomena. For this reason, qualitative studies should not be set aside. Otherwise, it would be easy to overlook many terminological interactions and negotiations. In short, the challenge is to create a corpus terminology in consonance with the uses of communities and sub-communities, and the terminological agreements that are negotiated within these discourses. The question then arises as which models should be used for the description of terminological, synchronic and diachronic variation.

Furthermore, there are other important challenges. In the world today, terminological resources must be flexible and modular. They should also be based on well-differentiated corpora that reflect how terms are used and the different discursive parameters that engage them. However, describing variation and being able to integrate it in resources are not quite the same thing. It is a question of collecting data and then being able to represent it. This involves considerably more than merely listing the variants related to a given community.

Terms affect the relational management of situations. Consequently, their meaning goes beyond the definition or the terminology record. The problem is how to integrate such information into a terminological resource. In a dictionary on nuclear energy, for example, the question is whether important issues can be brought out only by mentioning the conditions of production, circulation, and reception of certain term. It is necessary to remain faithful to words while contributing to what Rey (1979) called “le rôle thérapeutique du discours et du savoir”.¹⁶ This means not only clarifying meaning, but also usage. More broadly, this raises the question of orthonymy, which indicates that there is only one right word in a certain type of interaction, and accordingly, terms should respond to both sociolinguistic and cognitive needs.

The history of science abounds with examples of ideas and concepts that emerged because of social needs and contexts. One of the objectives of Socioterminology was to stake out a place for Terminology at the heart of Applied Linguistics, or *Situated Linguistics* as it is now often called (Condamines and Narcy-Combes 2015). According to Rey (1979, 24):

16. The therapeutic role of discourse and knowledge (our translation).

Une vue superficielle assigne à la terminologie l'étude des systèmes lexicaux thématiques et organisés des sciences et des techniques, la lexicologie s'occupant des sous-systèmes lexicaux du langage courant. Cette partition est inopérante : en fait s'agissant d'ensembles de mots ou d'unités lexicales et d'unités codées plus grandes que le mot, s'agissant de signes ou de signifiés, cette étude relève toujours de la lexicologie, de la linguistique.¹⁷

Instead of the dissemination of Socioterminology, we should be discussing the *infusion* of its concepts and methods, in the sense of Thibaudet (1936, 53) who spoken of “infusion de l'aventure et de l'inattendu dans la vie intérieure”¹⁸ or of de Carné (1835) with regard to the French Revolution: “l'humanité avait reçu à cette date je ne sais quelle infusion d'idées nouvelles”.¹⁹ Without a doubt, *infusion* is Socioterminology's greatest achievement.

Acknowledgements

We would like to thank Hazel Duncan for having translated this paper from French into English and John Humbley for his thorough re-reading of this article.

17. A superficial view assigns terminology to the study of thematic, organised lexical systems for science and technology and lexicology to lexical sub-systems in everyday language. This categorisation does not work. By examining sets of words, lexical and coded units larger than words, and signs and signifiers, terminology is firmly grounded in lexicology and linguistics (our translation).

18. The infusion of adventure and the unexpected into our inner life (our translation).

19. At that time humanity received I do not know what infusion of new ideas (our translation).

Cultural Terminology

An introduction to theory and method

Marcel Diki-Kidiri

Académie Africaine des Langues

The aim of Cultural Terminology is not the international harmonization of multilingual terminologies but the appropriation of knowledge and know-how by a specific speaker community rooted in its culture. Cultural Terminology puts the human being at the center of the terminological process. The human being, both as an individual and a community, is a cultural being who constantly creates culture through a mechanism of appropriation of the new as they learn and explore their environment. An in-depth analysis of this mechanism reveals the concept as an idealized mental representation, culturally motivated in its perception. Each concept can be perceived from several viewpoints called percepts. A precise methodology is outlined for terminology works and finally some examples of application are mentioned.

Keywords: Cultural Terminology, appropriation, culture, language planning, concept, percept, denomination

1. Introduction: Origin and purpose of Cultural Terminology

1.1 A Terminology for development

We have come to Terminology to instrumentalize African languages, with a view to making them suitable for use as tools for the transfer of knowledge and modern technologies appropriate for the development of African societies (Diki-Kidiri 1999b). Our approach therefore has as its ultimate goal the appropriation by a society of new knowledge, typically one that comes from a foreign culture. This implies the integration of the new into the receiver's traditional culture.

We also had to consider the appropriation of knowledge by individuals, especially as it is through the training of individuals that societies are transformed. Understanding the foundations of Cultural Terminology should lead us to the development of an optimized working method to ensure the successful appropriation

of new knowledge. As the Wüsterian General Theory of Terminology (Wüster 1976; Humbley, this volume) focusses on the international harmonization of specialized terms in different languages, it does not address the issue of the appropriation of new terms in any language, let alone a given culture, with the understanding that anything cultural is a disturbance for the clear communication of scientific and technical thought. These theoretical assumptions have been proven at least insufficient by four socially based new approaches to Terminology, namely Socioterminology (Gaudin 2005), Sociocognitive Terminology (Temmerman 2000), The Communicative Theory of Terminology (Cabr  2000), and Cultural Terminology (Diki-Kidiri 2008). These four trends re-introduce a human dimension in Terminology Theory. Socioterminology acknowledges the existence of multiple technolects based on the social status of workers within the same company. Sociocognitive Terminology studies the mental mechanism whereby we can understand specialized discourses. The Communicative Theory of Terminology focusses on the multidimensionality of the term as an object of communication.

Although these approaches share many aspects with Cultural Terminology, none of them addresses the appropriation of new knowledge by a given community of speakers. The point of view of Cultural Terminology is that terms are not just labels for concepts. Since both are produced within a culture, new terms must be culturally integrated to be or become real parts of the receiving society's knowledge. Hence, the appropriation of the new is *the* central issue here and the ultimate objective of Cultural Terminology. It is thus necessary to analyze the dynamic process of appropriation whereby people apprehend any new object, learn to know it, and give it a localized name, so that it becomes part of their culture.

1.2 The dynamic process of appropriation

Throughout their life, an individual or a community accumulates life experiences that become a body of knowledge used as a reference grid to apprehend everything new in their life. Each new reality is intuitively categorized in a prototypical way by comparison with the most similar archetype in this reference grid. If the categorization is totally satisfactory, the new reality is identified as 'already known' and retrieved. If the categorization is not entirely satisfactory, this dissatisfaction is often, but not always, reflected in the subsequent denomination. This is revealed by names such as: *false quinquelobed*, *false papaya*, *Cayenne cherry*, *island apricot*, *guinea pig*¹ etc., given by European travelers to new species encountered in Africa

1. In this case, it is neither a pig nor an animal of Guinea, since it is a small rodent from South America.

and America. Finally, if categorization is impossible, the new reality is classified as 'unknown', and is given an assumed name or a name inspired by one of its most salient features. This new reality will then occupy a place in the knowledge base of the individual or the society that has appropriated it. It can then serve as an archetype for the appropriation of other new realities that can be assimilated or brought back to it. As a result, each new appropriation increases the recipient's knowledge base and therefore his or her competence, while at the same time transforming his/her vision of things and his/her culture.

1.3 From reconceptualization to naming

Culture, both individual and collective, conditions the way people perceive and categorize reality. The same reality is perceived differently by people of different cultures (Roulon-Doko 1996; Diki-Kidiri 1999a). Each person's perception leads him or her to *reconceptualize* the object of knowledge in a different way to appropriate it. The resulting denomination is more a conventional instrument for intellectually grasping the object as it is perceived rather than a label that would serve the function of faithfully reflecting the structure or ontological nature of the object, except for some very narrow fields such as chemical formulae or certain mathematical concepts.

Cognitive science teaches us that a concept is already a construct of the mind that allows us to organize classes of objects in a prototypical way (Lakoff 1990). However, the study of naming makes it possible to highlight the fact that naming is very often motivated by *a cultural perception of the concept*, and not by its ontological structure. This cultural perception of the concept which is the result of a reconceptualization of the object is known as *percept* in Cultural Terminology. The chain of relations from the signifier to the object is the following: Object $\Leftrightarrow$ object classes $\Leftrightarrow$ concept $\Leftrightarrow$ percept(s) $\Leftrightarrow$ signifier(s). This chain includes the structure of the term which, as a linguistic sign, has at least three components: the signifier, the percept, and the concept. Conceptual variation is considered at the level of the synchronic diversity of percepts for the same concept and at the level of the diachronic evolution of concepts, whereas denominative variation is studied at the level of the signifiers that ultimately refer to the same concept. In Wüster's General Theory of Terminology, all these variations are ruled out, but not in Cultural Terminology.

2. Basic principles of Cultural Terminology

This section sets out the initial principles that are fundamental to understand the rationale behind the cultural approach to Terminology. Within this approach, Terminology is not a discipline that just stores information about terms in a specialized lexicon, but rather the study of terms as the cultural expression of specialized knowledge in any natural language belonging to a community of speakers. Therefore, human dimensions are fundamental to Terminology. Although these dimensions may not initially seem to be central to Terminology as it is usually addressed, they are the mechanism whereby human beings integrate in their culture a new experience or knowledge that will make them grow. These dimensions are the foundation and the justification of the cultural approach to Terminology. The cultural integration of new knowledge is based on the following: historicity, base of experiences and knowledge, memory, appropriation of the new and cultural growth.

2.1 Historicity

Every human being has a personal and unique history of their life which we call their *historicity* and whereby they build their personal culture made of their habits and knowledge. Similarly, each human community is formed ('born') at a particular moment in History and experiences a unique evolution, a historicity, made up of happy and unhappy events experienced collectively, shared activities, creations, movements of thought, innovations, dramatic situations, etc. All these things weave the particular History of any community. Historicity explains the diversity of cultures and the same time it shows how culture is a holistic concept that is present in any aspect of our lives.

2.2 Base of experiences and knowledges

Human beings accumulate an immeasurable quantity of experiences during their lives that constitute their knowledge base, their personal archives, and their memory. In the same way, all the facts that weave the historicity of the life of a community accumulate in the collective memory into a gigantic base of experiences and knowledge, which are accessible to the members of the community. The variety of technologies and cultures that human communities produce is as rich as it is immeasurable. Over time, communities evolve and so do their cultures. Exchanges between communities and the transfer of expertise and skills to new generations are important factors in the reassessment of values.

2.3 Memory

Of all the higher faculties of the mind, memory is the one that enables human beings to progress in the acquisition of knowledge. Memory enables the basis of accumulated experiences and knowledge to be exploited to enlighten judgment and thus behavior. Although complex in its components and functioning, memory ensures the encoding, storage, and retrieval of information perceived by the senses, which then informs judgment and consequently behavior (Taconnat 2012). Memory helps people to avoid making the same mistakes repeatedly because each experience teaches them a lesson and makes them more knowledgeable.

Regarding a community, it exists only if its members are aware of it and claim it. Collective memory maintains this awareness through the teaching of history, monuments, works of art, literature, and the perpetuation of traditions of all kinds. Collective memory is the storehouse for all the traumas suffered by the community (war, oppression, and persecution) as well as all the shared happy events. The commemoration of these events and the construction of symbols and referential values that are shared by the whole community facilitate communication between its members and maintain their internal cohesion and identity (Licata and Mercy 2015). No matter how it works, "...the collective memory would play a predominant role in the definition of a group's identity through the narration and sharing of their history but also in the justification of past, present and future actions of the group" (Wikipedia, 'Collective Memory'). Memory is the link between the perception of the new and the retrieval of the right archetype concept to identify it.

2.4 Appropriation of the new

The new can be an object, technology, discovery, fashion, behavior, theoretical concept, or absolutely anything which is not part of the set of experiences and knowledge already familiar to the individual or the community. How it will be stored and retrieved in the knowledge base depends entirely on its categorization.

Regarding the whole community, the new may come from the outside as well as from the inside. When the new comes from the outside, usually from another community where it is already conceptualized and well-integrated, the receiving community behaves towards it as an individual. It will perceive the new object, compare it with what it already knows to categorize it as accurately as possible, and finally name it to integrate it into its base of experiences and knowledge. This procedure is part of the dynamics of social change of which Rivière says that, in a nonconformist group, "any innovation is received there because of its novelty, while elsewhere one defies it because of its very novelty, because it risks disturbing the

traditional order” (Rivière 1978, 104). Obviously, the greater the cultural distance between the two communities, the greater the cultural clash wave that accompanies the appropriation of the new. What we call *cultural clash wave* is the set of transformations necessary for the new to be integrated into the culture of the receiving community. These transformations concern the reconceptualization of the new, the reformulation of its expression, and the overcoming of prejudices within the receiving community that may hinder or even prevent its appropriation. Evidence of this can be found in many studies of loanwords (e.g., Huttin 2015; Uffman 2015).

When the new comes from within the community itself, it generally originates in a quite small sector of it. The source can be a single person (a star who launches a fashion, an inventor, etc.) or a micro-community (a company, team of researchers, etc.) and the issue is to examine how it will spread within the rest of the community and become part of the community’s culture.

The use of metasemia (metonymy and metaphor) as a naming process in advanced or highly specialized disciplines has been commonly observed (Diki-Kidiri 1996; Lakoff and Johnson [1980] 2003). Metasemia consists in comparing the unknown entity with something that is already known, in order to better understand, explain, and know about it. It is the same process that is carried out when one appropriates something new from the outside. However, here, the cultural clash wave that always accompanies the appropriation of the new is significantly attenuated by the fact that this novelty originates from part of the same large cultural community, and not from another exotic one. The movement here is from an individual or specialist culture to a common culture in which the former is already immersed. The cultural clash wave therefore still exists, but its amplitude here is so small that it may be negligible.

2.5 Growth of the culture

Whatever the outcome of a new experience, it will reinforce or correct previous certainties, and allow a better knowledge of the environment or the outside world. The base of experiences and knowledge will necessarily be enriched by this new input. Accordingly, human beings, as individuals, may eventually decide to modify their convictions and/or behavior. For example, they may finally decide to stop smoking after the onset of a successfully treated throat cancer. Their personal culture will have radically changed because they now pay more attention to their health. For a community, the growth of the culture is like an in-and-out breathing movement. The inward phase goes from the identification of the new up to its incorporation into the community’s culture. Then, the new reality gradually loses its novelty, and

eventually becomes an archetype that can be used in the outward phase to help categorize other new realities for another appropriation. The culture is thus necessarily modified to a greater or lesser degree.

2.6 Culture

Culture is the totality of the experiences, productions, and knowledge generated by a human community living in the same space and in the same period. Hence, there is a diversity of cultures in both space and time, and a thick layering of culture that allows the various experiences and knowledge to settle in the archives of the collective memory of a community and in the individual memory of a person. These archives are shared symbolic references through which members of the same cultural community can understand each other when they communicate. In fact, words, gestures, behaviors, situations, and everything is more accurately interpreted and understood when people share the same symbolic references. Otherwise, a great deal of explanation is required to dispel incomprehension and misunderstanding.

From a diachronic point of view, culture is a historicity, in the above meaning of a 'particular history specific to' an individual or a community, which helps to shape their personality, especially in regard to their habits or customs and their relations with others. This aspect of culture is one of the foundations of the diversity of modes of expression and communication. Culture is thus like a microcosm which may appear strange from the outside, but which is completely coherent from the inside, since it shapes the totality of a person's relationship with whatever exists and thus governs their vision of the world, which is the way they perceive things. The next section will show that the way people perceive things and name them, and how they conceive them is culturally bound. Hence, the linguistic sign does not appear completely arbitrary as Saussure states it. So, a better understanding of the Saussurean statement is required to explain how the linguistic sign can be arbitrary on the one hand and yet culturally motivated on the other hand. Therefore, it is worth revisiting the complex relationship between concept, signified, and signifier in the light of culture to understand the Cultural Terminology approach.

3. Conceptualization, naming, and cultural perception

3.1 Meaning and concept in denomination²

The linguistic sign is arbitrary (Saussure [1916, 1959] 1975). Revisiting the famous Saussurean assertion that essentially, if not solely, concerns the relationship between the signifier and the signified, we shall focus on the difference between a linguistic sign and a concept.

Data from various African and European languages and cultures show that although the notions of 'concept' and 'signified' can both evoke the same cultural products of the human mind, the concept seems to express more objectivity and thus more universality in the representation of things, whereas the signified seems to be more dependent on perceptions particular to each culture (cf. Section 3.3.3. below).

However, just as the signified can change from one culture to another, it can also change in time and space within the same culture. For a given object, the concept is the essential idea, the principle, or even the archetype, whereas the signified is the angle from which it is viewed. This undoubtedly implies the existence of other angles of view.

The selection of an angle of view or viewpoint, of a perception or signified, depends on habits, analogies, and strategies of apprehension that are largely conditioned by the awareness of previous experiences, traces of which can be found in the memory of words (Delaveau 2007). Consequently, the relationship between the concept and the signified always includes a motivation that prompts and guides the choice of a signifier. The linguistic sign thus produced is not totally arbitrary. What is arbitrary is only the relationship between the sounds shaping the signifier and the signified. Even onomatopoeias describing the same fall of a heavy object in water differ from one language to another, such as *plouf* in French and *splash* in English.

Regarding the terminological instrumentalization of African languages for the expression of modern realities that did not exist or were not prominent in the context of traditional African cultures, an important issue is the complexity of the specialized knowledge unit that is the term. Naming new realities in African languages imposes the following constraints:

2. This presentation was the subject of a paper at the *V^{es} journées scientifiques du réseau Lexicologie Terminologie et Traduction (LTT)* [Vth Scientific Days of the Lexicology, Terminology and Translation (LTT) network] in Tunis (Diki-Kidiri 1999c). The present version has been extensively reviewed and extended for our recommended collective work published by Karthala, Paris (Diki-Kidiri 2008).

- a. The new realities to be named exist independently of African culture, since they were mostly created in the Western world. We must ask ourselves what their nature is and how these new realities are named in the Western cultures. Are their denominations accurate because they are valid designations of things that were objectively revealed by science? If not, why was the signifier chosen? Answering these questions, we hope to gain a better understanding of how best to name new realities in our respective languages.
- b. There are also ancient realities in Africa which have until now not been sufficiently studied in traditional African cultures. After encountering the West, these realities have become important for economic development, e.g., *cotton pests* (Tourneux 2006). Three types of observation are thus required: (i) the intrinsic nature of these realities; (ii) Western denominations; (iii) traditional African denominations. These observations reveal the extent to which the Western denominations are motivated by reality *per se*, and therefore show whether the existing African denominations are still operational, and hence recoverable. This evaluation should guide terminological renewal. However, only if we know what a 'good' denomination is, will we be able to develop 'good' terminology.
- c. The terms thus produced are linguistic units that will become an integral part of a natural language, and are intended to be used for communication both in the context of professional activities (Tourneux, 2006) and in the transfer of knowledge such as in schooling, literacy, training (Belloncle 1984). The term is therefore a linguistic sign that permits us to express and communicate specialized knowledge in a natural language.

If the term is indeed a Saussurean 'linguistic sign', the binary signifier/signified relationship alone is insufficient to describe the term in all its complexity. To delimit the field of Terminology as a scientific discipline, Cabré (1999c) proposes that terms have a linguistic dimension, a social dimension, and a pragmatic dimension. Yet, the linguistic dimension has three axes of description: signifier, signified (or percept), and concept.

The signifier refers to the morphology of the term, namely, word formation rules, homonymy, synonymy, denominative variation, etc. 'Well-formed' terms are easier to implement, provided that other aspects (e.g., syntactic and semantic) are also optimal.

The signified (or percept) refers to the construction of meaning which often implies a reconceptualization of the object to be named, based on cultural perceptions. The popular methods of meaning construction, such as metaphor and metonymy, often lead to polysemy, which must be addressed.

Finally, the concept deals with questions related to cognition and the nature of knowledge, its objectivity and its true or probable universality (Temmerman 2000). Several cognitive research studies (Lakoff 1990; Diki-Kidiri. 2008) amply demonstrate the constructed nature of concepts, including those that seem to be the most ‘independent of all subjectivity’. The scope of a denomination can be specified by exploring the nature of the relations between a concept and the various sets of objects it is supposed to represent (e.g., Goldstone, Kersten and Carvalho 2012; Wisniewski 2002). Finally, the distinction between lexical units in general language and terms in a specialized domain should be based on this triple articulation of the term that binds the signifier, the signified (or percept), and the concept. This is indeed new compared to the usual interpretation of the Saussurean term “signified” equated with concept. This differentiation is justified by our own research. It is therefore not surprising that there are no other references than our own publications (Diki-Kidiri 2008).

3.2 Concept and object class

In the case of a manufactured object or artefact, such as a house, a computer, bicycle, etc., the object was evidently designed before being made. This conception can be translated into a series of reflections, studies, calculations, etc. leading to an ideal mental representation of what the object is supposed to be from the point of view of its inventor. It is this mental representation that corresponds to the ‘concept’ of the object because it expresses the essential idea, principle, or archetype. This concept can be reduced to a written form called a ‘blueprint’ or ‘project plan’. Each object created with this plan is a concrete instance of the concept. In addition to the fact that all these objects share the essential attributes of the concept, each one of them may have its own personalized or localized attributes, which may justify a fractal structuring into object classes (Figure 1). This variation often begins at the conceptualization level and leads to a family of concepts. The Larousse Dictionary defines a concept as follows:

Idée générale et abstraite que se fait l’esprit humain d’un objet de pensée concret ou abstrait, et qui lui permet de rattacher à ce même objet les diverses perceptions qu’il en a, et d’en organiser les connaissances.³ (Larousse 2020)

This definition fully matches with our observations. Figure 1 depicts the fractal relationship between the concept with any specific object to which it refers. Object

-
3. General and abstract idea that the human mind has of a concrete or abstract object of thought, and which allows it to relate to this same object the various perceptions it has of it, and to organize their knowledge (my translation).

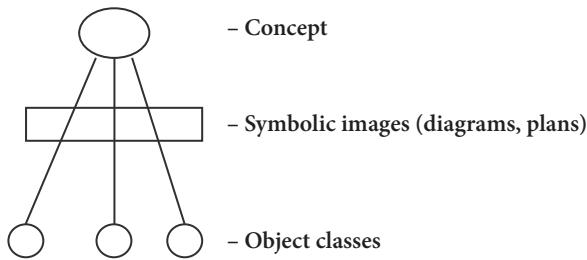


Figure 1. Concept to object class hierarchy structure

classes refer to the structure of the conceptual field based on common secondary distinctive features that define subcategories of objects. A percept is the perception, by a speaker or a community of speakers, of features salient to them because of their cultural background, and which motivate a denomination (the signifier). Several percepts can be found for the same concept or class of objects depending on the cultural perceptions of speakers.

Because the archetypal concept is at the basis of the production of the objects thus conceived, we are tempted to consider it as the expression of their intrinsic nature, independently of all contingencies. It is probably more accurate to consider the concepts giving rise to technological inventions as representations of the archetypes of these inventions. The archetype expresses the inventor's vision or the 'archetypal concept'.

When, for multiple reasons (e.g., use, evolution, cultural distance, etc.) technological products are perceived or presented differently from the vision of their inventor (i.e., their archetypal concept), we say that they have undergone a reconceptualization because of new perceptions, as often occurs in Biology.

When it comes to characterizing natural objects not created by human beings: animals, plants, spirits,⁴ stars, etc., the main issue is categorization, as very well shown by all research on cognition, prototypicality, etc. Typically, when an unusual animal is found, the more encyclopedic knowledge gathered about it, the greater the precision of its categorization. Thus, the whale was first classified as a fish before joining the category of mammals. Similarly, amphisbaenas were classified as snakes, then as earthworms before being assigned a specific category for them alone (Jared, Burman and Léon 1997). Biology, like all sciences, does not provide us with any guarantee of objective classes of natural objects which are independent of human perception.

Since human beings did not create the whale, their concept of it can only be constructed deductively by accumulating knowledge about it through observation.

4. In most African cultures and languages, spirits are regarded as natural entities.

This creates a schematic mental image of the whale comparable to the one created by the inventor of an artefact. This mental image incorporates enough of the whale's typical features for any whale to be recognized when encountered. The concept of 'whale' underlying this symbolic reconstruction of the mental image of the whale is no different from an archetypal concept, despite the inversion of the path (objects to concept here, concept to objects in the case of artifacts). So, in both cases, there are always three components: archetypal concept, symbolic diagrams, and classes of objects.

The question that arises here is whether the categories established in our mind, either because of observation or mental activity (the inventor's flash of genius) really give us access to the knowledge of things as they are in the world outside of our mind. Terminologists are interested in knowing more about the mind in so far as it enables them to describe, for example in a terminographic definition, the salient features of a specialized knowledge object, whether it is an artefact or a natural species. A good definition facilitates the resolution of the naming issue. However, this naming issue should not be dealt with at the concept level, but rather at the level of the interaction between the signifier and the signified organized by cultural perceptions.

3.3 Denomination and cultural perceptions

3.3.1 *Diversity in the observation of reality*

Research largely supports the hypothesis that human beings can obtain knowledge of the real world only through culturally conditioned mental representations (Diki-Kidiri 1999a–b). The division of reality often varies from one culture to another, giving rise to concepts specific to each culture. For example, in several African languages sharing the same cultural area, colors are generally classified into three categories which could be referred to in English as 'dark', 'light', and 'bright'. In contrast, in European cultures, the same colors are categorized as a succession of individual shades as shown in the rainbow color scheme. This is a difference in the conceptualization of colors, not in their psycho-physiological perception. Terminology research shows that the same technological product conceived in one culture and imported into another can be integrated into the latter only through a process of reconceptualization which is a part of the phenomenon of appropriation of the new (See Section 2.4).

To give just one example out of many, the choice of the *logiciel* and *matériel*, as the French translation of 'software' and 'hardware' respectively required a long process of reconceptualization during which thirteen other pairs of candidates were discarded. The English language classifies the components of computer resources into 'soft' and 'hard'. The soft/hard opposition is already, in itself, a cultural

classification. Neither magnetic tapes nor floppy disks, which were used in the 1980s to save programs and data, were parts of the ‘soft’ set, whereas they were opposable to the ‘hard disk’. French speakers have had to reconceptualize things based on their way of thinking, whose Cartesian symbolic reference is highlighted in collective memory. ‘Software’ is then categorized as everything that generates and governs the ‘logical’ flow of operations performed by the computer. As the latter is easily categorized as *matériel* [equipment], the creation of the neologism *logiciel* [software] from the word *logique* [logic] and the suffix *-iel*, as in *matériel*, was obvious, because the new term was coherent with the French language.

3.3.2 Polysemy and prototypical organization of meaning

The French word *souris* [mouse] prototypically designates a ‘quadruped of the rodent family, belonging to the genus *rat*’ (Littré 2020). This means that this is the primary or first meaning of *souris* [mouse] in the French language. It is also prototypical because this is the meaning most people think of when they hear or see this word. However, the word *souris* [mouse] has several other secondary or figurative meanings in the French language such as the following:

- a man who is very afraid or feels great embarrassment.
- a silver-grey color.
- a fleshy muscle that is attached at one end to the sleeve of the leg.
- the hole in the hand formed by the thumb and the forefinger when their ends join.

In Computer Science, as we know, the word refers to an electromechanical or electro-optical device for pointing and grasping. So, we have six different meanings for the same word *souris* [mouse]. This is indeed a case of polysemy. Since polysemy as a linguistic phenomenon has been studied in some depth, it will not be revisited here. We will simply highlight a few points.

The signifier *souris* [mouse] (S) has six meanings in French, one meaning that is central (S_0 = ‘rodent sp.’) and five that are non-central (S_1 = ‘shy man sp.’, S_2 = ‘gray sp.’, S_3 = ‘muscle sp.’, S_4 = ‘hole sp.’, S_5 = ‘pointer sp.’). These indices can be numbered from 0 to 5, one for each of the six senses without any hierarchical structure implied. It is easy, in fact, to explain these different meanings of the word *souris* [mouse], notably by analogy, metonymy, metaphor, etc. (Lakoff and Johnson [1980] 2003) However, the question here is to show how naming is distinct from categorization.

From the point of view of categorization, only the small animal designated by the word *souris* [mouse] can be classified as a ‘rodent’. The other signifiers (S_1 to S_5), however, refer to other conceptual categories. Even frightened, a man is not ontologically a rodent, nor is a mobile muscle, etc. However, in the French culture, the frightened or embarrassed man remains frozen or runs frantically in

all directions. He has a behavior perceived by the French as similar to a frightened rodent's behavior. The reduced space between the thumb and index finger evokes to them a mouse hole. The silver-grey color being dominant in mice, the synecdoche is understandable. For the muscle as for the electronic device, it is the similarity of shape that is sufficient to allow the comparison. All these derived meanings have no feature in common. However, each one shares one feature with the archetypal concept of the rodent.

In Figure 2, the propagation of the lines from the central S_0 to the five other is radial. The centrality of S_0 , which designates the archetypal concept C , may vary over time. For a generation of children living in a large western city like Paris and who grew up with video games, *souris* [mouse] primarily evokes the computer accessory. This meaning (S_5) supersedes and relegates to the background the original meaning of 'rodent'. This means that, even in polysemy, the centrality of a meaning, of a signified is directly dependent on the social environment to which the majority of the respondents belong. And if we consider the evolution of the meaning of a word over time, it is not uncommon to find that a new meaning has completely supplanted an older one that has been totally forgotten, except perhaps in the encyclopedias. The central meaning of *keyboard* today is no longer that of "a board on which keys are hung" as it was in the Middle-Ages and the French word *clavicule* [collarbone] no longer means "small key". What was a metaphor, and thus a non-central signified, has evolved to become the primary meaning, and even the sole meaning.

Nevertheless, these forgotten meanings are only overshadowed. It suffices to place a narrative in a historical context for them to resurface. Such meanings can

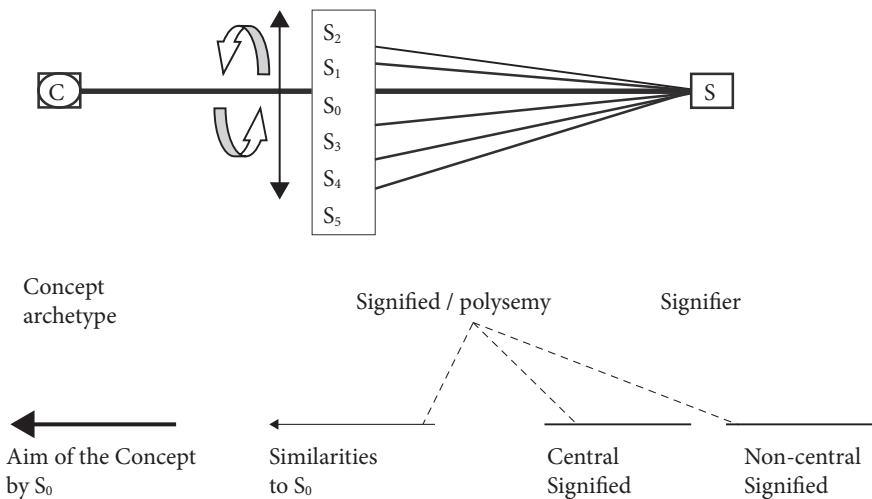


Figure 2. The linguistic structure of the term, from signifier to concept

always be reactivated, even without this type of context. This is what occurs when a word is deliberately used in its etymological sense. Thus, in the same language, meanings that are polysemically linked to the same signifier can be organized both in space and in time and according to relatively homogeneous social environments. It follows that a non-central meaning in a given environment, at a given time, and in a given place, can occupy a central position in another temporal and situational environment. This is indicated in Figure 2 by the rotating arrows surrounding the central meaning.

3.3.3 *Concepts and the signified or percepts*

The distinction between concept, understood as a cognitive structure of categorization, and signified, envisaged as the place of cultural perceptions becomes even more evident when we compare the denomination of the same artifact in several languages. The bicycle is called:

- i. *gbâzâbângâ* ‘rubber wheels’ in *Sanngo* (Central African Republic),
- ii. *nàgàsó*⁵ ‘iron horse’ in *Bambara* (Mali),
- iii. *magu-mákwanganya*⁶ ‘four feet’ in *Lilikó* (Democratic Republic of the Congo).

These different names bear witness to the diversity in the perception of the bicycle by communities of different languages and cultures, with different pasts. Central Africans were familiar with the wheel, and moreover had been subjected to forced labor to harvest vegetable rubber since the initial colonization of their country (in 1910 or so). The rubber wheels (tyres) of the bicycle therefore particularly caught their attention and motivated their choice of the name of this vehicle. In contrast, the *Bambara*, who were more familiar with horses, perceived a functional resemblance between this animal and the bicycle. Both are mounted astride and allow humans to move faster than on foot. However, only the latter feature (moving faster than on foot) caught the attention of the *Boliko* of the Democratic Republic of the Congo where king’s messengers usually run to deliver messages. They subsequently see the bicycle as ‘a vehicle that multiplies your feet’ and allowed them to go twice as fast.

In all cases, these names are not intended to account for the schematic structure (archetypical concept) of the bicycle. Each of the signified corresponds to different points of view or perspectives on the object, motivated by and dependent on the cultural past or ‘historicity’ of each community. The signified is therefore not the equivalent of the concept, but only a pointer to the concept, an anchoring point

5. Information confirmed by Gérard Galtier in a personal communication.

6. Information given by Edema Atibakwa Baboya in a personal communication.

which allows us to globally grasp the concept without need to reconstitute all its structural elements (see Figure 3). For this reason, we call it a ‘percept’ though this term has a different definition in Cognitive Psychology.

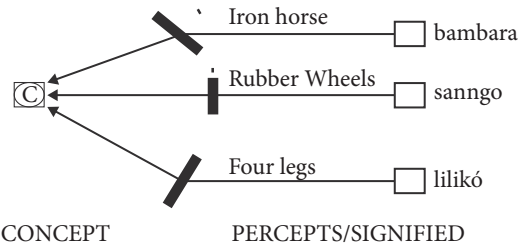


Figure 3. Diversity of percepts for one concept

In the signified/signifier Saussurean dichotomy,⁷ the term *signified* globally designates all the semantic content related to the signifier. It is still very commonly used as a synonym for *concept*, or even *notion*. In that broad interpretation, it covers both ‘percept’ and ‘concept’. However, since the linguistic sign is a minimum unit of meaning, the signified that composes it should be understood as a minimum significant value. This makes it possible to use an entire linguistic sign (with its signifier and its signified) to express, in a motivated way, a meaning or concept that may have nothing to do with the minimum value of the signified. This restricted interpretation of the term ‘signified’ allows us to state that the percept is a signified within the framework of the terminological unity that is the term.⁸

4. Methodology in Cultural Terminology

4.1 The socio-professional framework

The application of theoretical principles always requires an orderly working method with objectives and pragmatic conditions of execution (Cabr  1999c; Diki-Kidiri 1998). The general framework of this application can be a large-scale language policy initiated by an authoritative body such as a government or a non-governmental organization. Therefore, when embarking on a project in Cultural Terminology, it is better to start by clearly delimiting the sociolinguistic and socio-professional

7. The Saussurean linguistic sign has two distinctive components with completely different features: the signifier and the signified, therefore it is a dichotomy.

8. See the definition of *term* given in paragraph c) in Section 3.1.

frameworks that fit the project. The sociolinguistic and socio-professional frameworks are determined by three dimensions: a scale, a sector, and a target (Diki-Kidiri. 2008). The subsequent terminological work is then conducted within this framework.

4.1.1 *The scale*

Scale is the vertical dimension of the sociolinguistic framework. Within this dimension there are at least four levels ordered from bottom up: local, regional, national, and international.

The local level is that of a local community: namely, agricultural cooperative, village association, rural or urban communes, private companies, etc. (Beloncle et al. 1987; Roulon-Doko 1998; Tourneux 2006). The results of the terminological activity here are primarily intended for local application, for example, the fixing and standardization of a technical vocabulary within a company, a district, a village cooperative, etc.

The regional level concerns an area covering an extended part of a state, such as a department, region, province, etc. The terminological activity aims to apply throughout the area, for example, to develop a regional lingua franca (Tourneux, Daïrou and Boubakary 2017).

The national level represents that of the highest administration in a national community (state or federal state). The terminological activity aims at the consolidation of a national language in all the activity sectors of the nation (Diki-Kidiri 1992; Morill 1997; Moser 1992).

Finally, the international level is the scale of international cooperation in Terminology. This is the case for vehicular cross-border languages, harmonization of working methods, or establishment of standards, etc. It is obvious that it will be necessary to define the levels of the scale depending on the realities of each field of activity relevant to each project, in particular the intermediate levels between the local and the international.

4.1.2 *The sector*

Determining the socio-professional sector on a horizontal dimension allows us to define the optimal size of the nomenclature. Indeed, it is not necessary to translate all medical terminology if it is only a question of training first responders. The nomenclature should contain only the vocabulary that is necessary. Once the sector has been defined, various tasks should be scheduled to address and fully cover this sector.

4.1.3 *The target*

The target is the result of the scale reported to the sector. It includes the immediate purpose of the terminology project, and where appropriate, the end users of the results of the work, who should always be strongly associated with the different

phases of the development of the project work. For example, if the terminology is developed to facilitate the technical training of agricultural workers on a farm, it is essential to involve these workers in the selection and production of the terms in their language (Tourneux 2006). The target will be different if the goal is to develop agricultural terminology for vocational training schools. In this case, it would be better to request the collaboration of teachers from at least one of these schools.

4.2 Data collection

4.2.1 *Establishing the source*

If the project objective is to establish for an African language a standardized terminology regarding a professional activity or a technical field which is already well documented in a European language, the first step is to select the source nomenclature. This can be done with the collaboration of a specialist in the field or using reference documentation pertaining to the sector. The important thing here is to collect the right documentation for the project.

4.2.2 *Establishing the target*

It may be necessary to first analyze the percept that emerges from the signifier the term in the source language and compare it with the concept as described in the definition of the term, in order to understand the process of cultural perception which led to the denomination in the source language. From there, we can easily search and find one or more percepts in the target culture which permit a reconceptualization of the concept in the target language. All proposals will be noted and discussed with specialists of the target culture and/or language in order to retain only one candidate per term if possible. This results in a provisional list of candidate terms in the target language. Sometimes, two candidates may be selected as synonyms.

4.2.3 *Sources of information and documentation*

In addition to the experts, other reliable sources of information can be used, such as museums, monographs, ethnoscience publications, etc. It is advisable to collect as much documentation as possible on the culture, technology, and general knowledge of the populations who speak the target language. The exploitation of these sources should lead to the identification of concepts already well established in the host culture, but which may be perfect or partial equivalents of concepts identified in the source nomenclature. Finally, lexical creation in accordance with the word-formation rules in the target language can provide fully acceptable equivalents for the source terms.

4.3 Establishing the equivalents

4.3.1 *Perfect equivalents and quasi-equivalents*

By comparing the two lists of the source and target terms, it is easy to identify all the source terms for which a suitable target term has been found. These can be seen as ‘perfect equivalents’. It always seems easy to establish equivalences when the concepts seem identical in the source language and the target language. However, even in this case, it is useful to take the time to reflect to ensure that the equivalences are appropriate. As is often the case, after going through the whole list and comparing the equivalents established out of context, it is frequently necessary to discard initial first choices.

Quasi-equivalents are terms which refer to partially equivalent concepts out of context. The target language candidate term has a signification scope that is wider or narrower than the one in the source language. It is necessary to decide to what extent this equivalence is acceptable. In addition, there may be several synonyms or quasi-synonyms in the target language that stand for the same source language concept, even after its reconceptualization in the target language culture. In this case, it is preferable to choose the term which has the fewest connotations, is easier to pronounce, and fits more easily into texts. If necessary, standardization should be established, while respecting professional practice in the field. From this perspective, quasi-synonyms, such as *plug*, *connect*, *reach*, and *join*, may be used in different semantic contexts. Accordingly, devices are ‘plugged in’, users are ‘connected’, a correspondent is ‘reached’, and a couple is ‘joined’ (in matrimony).

4.3.2 *Identification of concepts without equivalents*

When a concept has no equivalent in the target language, it is necessary to analyze the conceptual content of the term in the source language to properly circumscribe the unit of knowledge that it denotes, and then analyze the perception behind its denomination. This information is used to reconceptualize the perception of this content to find a name in consonance with the culture of the target language. To do this, we start by exploiting all the possibilities of lexical creation offered by the target language. Very often this solution produces satisfactory results. However, if no satisfactory output is obtained, the following procedures can be implemented to find a similar concept that can metaphorically be used to coin a new designation:

1. Exploring the culture of the target language,
2. Exploring the neighboring language varieties of the same cultural area,
3. Borrowing from other African languages from the same cultural area with frequent exchanges,
4. Borrowing from the source language and adapting the loan word to the target language structure. This last solution is to be used with great care and only when all the other solutions do not lead to a better result.

4.3.3 *Processing variants*

Establishing bi-univocity between a denomination and a definition is the ultimate aim of any terminological activity. This principle is necessary both to accurately express knowledge precision and facilitate its transmission. That is why we need a language with a stable and standardized technical vocabulary for professional use. However, it has been shown (Temmerman 2000; Gaudin 2005; Diki-Kidiri 2008) that in real life, denominational variation is a discursive, educational, and social necessity. Paradoxically, it contributes to the optimization of the expression and facilitates the transmission of knowledge, especially in teaching contexts. When establishing a specialized vocabulary, the principle of bi-univocity should be applied to the extent possible, when recording the variants which seem stable enough to constitute possible alternatives. Evolution in usage or knowledge can lead to a readjustment of the terminological standard established at a given time.

5. Conclusion

Cultural Terminology has been developed to empower speaker communities of less endowed languages who want to access new technologies and science while preserving their cultural identity. It is particularly recommended in a setting where what is new comes from the outside and must be appropriated by the recipient society. It has been successfully applied by Mensah (2008) and Ramavonirina (2011). It has also been used to create a Sanngo⁹ vocabulary for the legal terminology of the International Criminal Court to train court hearing interpreters (Diki-Kidiri 2014). By applying the methodology of Cultural Terminology, the Institute of Applied Linguistics of the University of Bangui (Central African Republic) has published a significant number of lexicons and technical handbooks in mathematics, law and administration, farming, elections and citizen's rights, Sanngo syntax, broadcasting, and office suite (Diki-Kidiri et al. 2005) to enrich Sanngo as a national language.

Cultural Terminology has influenced the work of many researchers (Kadima 2013; Baraké 2016). Rather than merely producing 'labels' for new concepts, Cultural Terminology requires an extensive and in-depth knowledge of the language and the culture of the community whose language is under development. Such knowledge is much better achieved through interdisciplinary collaboration between a linguist, a domain expert, and a culture specialist. Works such as those of Roulon-Doko (1996, 1998), Tourneux (2006, 2008), Tourneux, Daïrou and Bubakary (2017) that apply a similar methodology to ours, are of paramount importance for terminology work.

9. Alternative spellings: *Sango*, *Sängö*, *Sahngo*.

PART 4

**Textual Terminology, Terminology
and Lexical Semantics**

Textual Terminology

Origins, principles and new challenges

Anne Condamines and Aurélie Picton

Université Jean Jaurès / TIM-FTI, University of Geneva

This chapter deals with Textual Terminology, an approach that emerged in the early 1990s along with other approaches presented in this book. One of its main characteristics is that it arose from an encounter between related disciplines that focused on the use of text mining (Natural Language Processing) and modeling tools (Knowledge Engineering and Artificial Intelligence). These disciplines have worked together to consider the new needs related to terminology, and their evolution. The tools developed and the diversification of needs have resulted in significant methodological developments in Textual Terminology. This chapter traces these developments and discusses the original bases, principles, challenges and perspectives that constitute Textual Terminology.

Keywords: Textual Terminology, specialised corpora, methodological principles, challenges, distributional analysis, knowledge-rich contexts, applied needs

1. Introduction

This chapter is a reflection on Textual Terminology from its emergence in the 1990s to the new issues it faces today. The second section of the chapter identifies the main elements that contributed to its foundation in the 1990s. Special importance is placed on the tools and collaboration with research in Natural Language Processing (NLP), Knowledge Engineering (KE) and Artificial Intelligence (AI). The third part focuses on the methodology of Textual Terminology as well as the systematic framework for tool-assisted methods aimed at analysing specialised corpora. In particular, we discuss the place of users in the overall analysis and stress the need for collaboration at each stage of the process. The fourth part highlights the modalities and possibilities that Textual Terminology should encompass, in the light of the parallel evolution of corpora, needs and approaches in Textual Terminology, NLP, and KE/AI. This perspective enables us to better situate Textual Terminology in the context of current approaches and to propose research issues to be addressed in the short and medium term.

2. The origins of Textual Terminology

Textual Terminology arose from a combination of the following: (i) new needs in relation to Terminology; (ii) the development of Corpus Linguistics; and (iii) the encounter between Natural Language Processing and Knowledge Engineering/Artificial Intelligence.

2.1 New needs in relation to Terminology

Until the 1990s, terminology studies were mainly related to translation needs. However, because of the need to manage and analyse digital documents and resources with NLP tools, terminology came into the spotlight as a resource for managing electronic documentation. Tasks such as document storage, information search and retrieval, and even translation all required up-to-date terminologies adapted to real-world use.

The structure of databases had to be adapted to new tools, and their entries also needed optimisation. For example, in a study conducted in 1992 at Matra Marconi Space (an aerospace company, now Airbus Defence and Space), only 15% of the terms found in the documentation in French concerning test/simulation benches were recorded in the official terminology in the “Vocabulaire des sciences et techniques spatiales”.¹

The need was then no longer limited to the creation of terminology standards. What was required was the creation of repositories that corresponded to actual term usage and were usable by tools. It was thus necessary to base the structure of terminology repositories on the analysis of digital corpora that were beginning to be available in large quantities in companies.

However, in the early 1990s, most terminology studies were carried out within the framework of the General Theory of Terminology (GTT), developed in the 1930s by Eugen Wüster (see chapters by Candel and Humbley in this volume). Wüster was a Viennese engineer, who advocated a prescriptive view of terminology with the objective of guaranteeing non-ambiguous communication in science and technology and especially within companies, both in the same language as well as different ones. Hence, one of the cornerstones of the GTT was the “pious wish” of biunivocity (“ein frommer Wunsch”, Chapter 1 in this volume): one term <> one

1. Edited by the DGLFLF (Délégation Générale à la Langue Française et aux Langues de France, General Delegation for the French Language) <https://www.culture.gouv.fr/Sites-thematiques/Langue-francaise-et-langues-de-France/La-DGLFLF>. Accessed October 27 2020.

concept. Polysemy was thus not acceptable. Although Wüster was familiar with the work of linguists such as Saussure, he believed that it was necessary to separate terminology from discursive variations. One explanation for this is that Wüster was firmly convinced that concepts were the starting point for Terminology (Wüster 1985 [1979]). He acknowledged the link between terms and documents, as its reflection was anchored in information sciences (Campo 2012, 41–43). However, partly because automatic processing was not strongly developed yet, he did not question the systematic use of corpora as knowledge reservoirs. The onomasio-logical approach of the GTT was then harshly criticised by corpus linguists, especially when electronic corpora (internet and intranet among others) became widely available.

2.2 Corpus Linguistics and corpus-based Terminology

Corpus Linguistics began to appear before the 1990s. One of its main objectives was the construction of general language dictionaries. Specialised texts were thus excluded as the corpora had to represent the general language. According to Sinclair (2005, 14), “[a]nother factor that affects balance is the degree of specialisation of the text, because a specialised text in a general corpus can give the impression of imbalance”. However, the methods developed to study general corpora were subsequently used for specialised corpora: “[c]orpus-based terminology can best be characterised as a working method which explores a collection of domain-specific language material (corpus) to investigate terminological issues” (Gamper and Stock 1998, 149). This is the case for definitional contexts or conceptual relationship markers (Auger and Barrière 2008; L’Homme and Marshman 2006). And indeed, part of Corpus Linguistics was quickly devoted to the use of specialised corpora to build terminologies. The use of corpora led to a rapprochement between terminologists and linguists. For terminologists, it resulted in the discovery of term variation and the need to propose theories that were more in line with these variations. Sager (1990, 58) writes: “[t]he increasing tendency to analyse terminology in its communicative, i.e., linguistic context, leads to a number of new theoretical assumptions and also new methods of compilation and representation”.

One of the main aspects investigated concerned the social dimension of Terminology (see chapter by Gaudin and Delavigne in this volume). The study of specialised texts means that it is necessary to adopt a situated vision of how language functions, in other words, to consider text production and its context. For this reason, Terminology could no longer be based on a purely onomasio-logical approach which mainly focused on the creation of norms. Researchers throughout the world began to criticize Wüster’s vision and decided to study real term usage,

(e.g., Condamines 1995; Gaudin 1993, 2003; Temmerman 2000). For example, Cabré (1999c, 114) argued that “Terminology, as both a discipline and an activity, but especially as a discipline, needs a new orientation which stresses its social and pragmatic aspects.”

In Textual Terminology, as for the French semanticist Rastier (1998b), it is neither the term nor the corpus that constitutes the most significant element but rather the text (see Chapter by Pecman and Kübler in this volume). It is mostly at this level that the communicative situation, the domain and especially genre should be considered. If texts are collected to form a study corpus, it is because they belong to genres that are relevant to the purpose of the study. In such an approach, the interpretation of language data needs to take the nature of the texts and the objective of the study, into consideration.

In addition to the crucial role of texts, the other defining characteristic of Textual Terminology is the use of NLP and Knowledge Engineering to analyse corpora and structure terminologies.

2.3 Terminology, NLP and Knowledge Engineering

The 1980s witnessed the approximation of terminology and knowledge (or Artificial Intelligence). In 1985, Nedobity wrote a paper entitled “Terminology and Artificial Intelligence”. In 1987, the first Terminology and Knowledge Engineering (TKE) conference took place at the University of Trier. In 1991, Skue and Meyer wrote an article entitled “Terminology and Knowledge Engineering: Exploring a Symbiotic Relationship”. In 1993, the Terminology and Artificial Intelligence group in France brought together computer scientists, linguists and terminologists, who played a major role in the development of Textual Terminology. These few examples show that, as early as the end of the 1980s, the close relationship between the issues of Terminology and those of Knowledge Engineering (or Artificial Intelligence) was perceived.

Clearly, these disciplines were all united by a shared interest in representing knowledge in the form of networks of terms/concepts. Knowledge engineers realised that they could use existing terminology databases and/or methods inspired by corpus-based terminology (Aussenac-Gilles, Bourigault and Condamines 1995; Aussenac-Gilles, Biébow and Szulman 2000b; Buitelaar and Cimiano 2008), and terminologists realised that they could use the knowledge representation modes of knowledge engineering to help build coherent and comprehensive resources (Meyer et al. 1992). In fact, the representation of knowledge in the form of a graph was perfectly adapted to natural language processing tools.

With the development of ontologies, the relationship with Terminology became even clearer, which led to the emergence of termino-ontological resources (Bourigault, Aussenac-Gilles and Charlet 2004) or termontology resources (Kerremans, Temmerman and De Baer 2008). In regard to Terminology, the evolution of databases enriched with insights from Knowledge Engineering tools gave rise to terminology knowledge bases. The main objective of this representation model is to replace, insofar as possible, a definition by a network of conceptual relationships (Condamines 2018c).

The need to systematically build a network of concepts from corpora led to the development of a significant number of studies and tools based on the notion of markers of conceptual relations (and knowledge-rich contexts (Meyer 2001; Marshman in this volume)). Based on this same objective, NLP and Terminology pursued their own research agendas. For NLP, it was a question of data extraction and retrieval though focused on the identification of conceptual relations (Bourigault, Jacquemin and L'Homme 2001). For Terminology, it was a question of seeing how to adapt lexicological studies on the construction of local grammars with a view to identifying conceptual relations in corpora. In relation to the phenomena observed in specialised corpora, various studies focused on the role of textual genre and the effectiveness of markers (Condamines 2002, 2008b; Marshman, L'Homme and Surtees 2008a).

3. The principles of Textual Terminology today

This section describes the methodological principles underlying the approach of Textual Terminology, the stages involved as well as the main protagonists.

3.1 Steps in a textual terminology analysis

In Textual Terminology, linguistic analysis typically takes place in six stages: (i) definition of the (theoretical or applied) purpose of the analysis; (ii) compilation of the corpus; (iii) selection of tools; (iv) use of tools; (v) validation of the results; (vi) integration of the results into linguistic paradigms.

3.1.1 *Purpose of the analysis*

The purpose of the analysis may stem from an extra-academic need or from scientific inquiry, often triggered by an extra-academic demand. In any case, especially when it concerns an extra-academic demand, the request needs to be 'translated' into a linguistic question. This makes it possible, on the one hand, to use knowledge

already acquired in Linguistics and, on the other hand, to integrate the results acquired into the chosen linguistic paradigm once the analysis has been completed. It is then a question of understanding how studies that are doubly situated (by the objective and by the nature of the corpus studied) contribute to the advancement of knowledge.

For example, a few years ago, to celebrate the 50th anniversary of its creation, the CNES (Centre National d'Études Spatiales, the French National Space Agency) asked us to carry out a research project on how space, as a body of knowledge, real or imaginary, was more or less consciously disseminated to the general public. We proposed to address this question from the perspective of determinologisation, i.e., the movement of terms from specialised language into general language (Meyer and Mackintosh 2000). This seemed a good indicator of the penetration of the domain-specific knowledge into general knowledge. A thesis studies this question – practically unexplored in Linguistics to date – in the domain of Particle Physics (Humbert-Droz 2021).

3.1.2 *Compilation of the corpus*

At the beginning of Corpus Linguistics, a corpus was intended to represent a (sample of) language, mainly with a view to building a general language dictionary or even a grammar. To ensure their representativeness, well-balanced corpora were built to report on a variety of communicative situations, but not specialised communication (see Section 1.2). Even though representativeness might seem easier to achieve in a specialised corpus, the notion of 'domain', does not really simplify the problem of corpus compilation. This notion is generally based on a top-down view. In other words, knowledge is organised *a priori* into domains within which discourse and terminology are assumed to be structured. This division is already problematic, as knowledge is not naturally discontinuous (Rogers 2013).

Moreover, this pre-established organisation is only partially relevant when it comes to considering a particular need. From the point of view of Textual Terminology, the corpus should be as representative as possible, but this representativeness also encompasses the objective of the study (Humbert-Droz, Picton and Condamines 2019). In the case of the previously mentioned determinologisation study, the question of representativeness also takes into consideration the types of texts through which terms are disseminated. We thus chose to build a corpus composed of four sub-corpora: (i) a corpus presenting projects carried out at CNES; (ii) a corpus of press releases issued by the CNES; (iii) a corpus of press releases containing space terms; and (iv) a corpus from the general press (Condamines and Picton 2012).

Another important point within Textual Terminology concerns the comparison of sub-corpora. Depending on the study to be conducted, the corpus is organised into sub-corpora, which are compared in a way that is relevant to the study. Corpora

can thus be organised based on time period, levels of expertise, disciplines, different departments within a company, etc.

One difficulty with building a specialised corpus concerns the availability of the texts and the difficulty of accessing data. There are various reasons for this, some of which are specific to the analysis of language for specific purposes. For instance, as Ghaddar and Langlais (2020, 3600) note, in an industrial context, “privacy issues and the commercial value of financial domain data” make it difficult to access data in specialised fields. Grabar et al. (2019) also discuss individual privacy and data protection issues that limit access to certain kinds of medical texts. Moreover, in order to systematically study the productions with the help of tools, texts must be in written and digital form. This requirement excludes, for example, handwritten notes, radio programs (unless they are transcribed), lectures, or even confidential publications. These constraints can result in rather small corpora and create methodological problems (see for instance Bowker and Pearson 2002).

3.1.3 *Selection of tools*

Tools analyse character strings and not words as semantic units. They quantify these strings, find similarities between them, and/or search for predefined strings in a corpus. The objective of NLP researchers is to propose results that could be associated with a certain meaning. In Textual Terminology, the tools are designed for a specific task (i.e., terminology and conceptual relation extractors). They may be more generic (e.g., taggers, concordancers, lexicometric, textometric tools, etc.). It is therefore necessary to decide which tools to choose, depending on the objective of the study.

3.1.4 *Use of tools*

Since Textual Terminology proposes to enter the corpus *via* terms, for whatever purpose, a term extractor should be the first tool used. Once the initial list of candidate terms has been established, the next step is to analyse the contexts in which they appear. There are two ways to study such contexts. The first way is to study the meaning based on the distribution of terms in the corpus (bottom-up analysis). The second way is to search term contexts for elements that are known to be relevant to the study (top-down analysis).

Bottom-up analysis: This type of analysis involves categorizing term contexts and then interpreting them (based on our competence as speakers and linguists). It is based on the well-known distributional approach developed in two currents, one inspired by Sociolinguistics (Firth 1957) and the other by a mathematical (and behaviourist) vision of language (Harris 1951). The main idea underlying this approach is that there is a correlation between distributional similarity and meaning similarity. It is thus possible to understand the meaning of a word or group of words by considering all its contexts.

Most analyses in Textual Terminology focus on corpora and the comparison of sub-corpora. In such cases, differences in distribution are associated with differences in the communicative situation, prior to the sub-corpus organisation. For example, if the corpus is organised in sub-corpora of experts vs. non-experts (e.g., to study popularisation as in Delavigne [2017] or Ledouble [2020]), a difference in term distribution, such as when the term is used in general language, can be interpreted as a sign of meaning variation.

Top down analysis: In the early days of Textual Terminology, when the main purpose was the construction of terminology networks, research focused on defining and specifying conceptual relation markers or lexico-syntactic elements to which an *a priori* interpretation is associated. The hypothesis is that these markers, associated with *knowledge-rich contexts* (Meyer 2001), are less likely to vary from one specialised corpus to another. The choice of appellation is very interesting as it introduces the notion of “wealth of knowledge”. This is not only relevant to conceptual relation markers but also to linguistic units of different types, related to different types of knowledge. This means that for each analysis, elements that are directly related to the objective of the study can be defined *a priori*. The purpose here then is not to build a relational network, but rather to track the presence of elements that indicate a shift in meaning, variation, etc. that speakers are aware of.

For example, if the analysis focuses on the evolution of a term, word, or phrase over a period of time, one can find markers such as, *before*, *[concept name2] formerly called [concept name1]* (Picton 2009, 2014b), that “mark” the presence of relevant information regarding the evolution of a domain or its terminology.

3.1.5 Validation of the results

In Textual Terminology, the results obtained must be validated at many levels. Firstly, results must be acceptable to peer analysts. This implies that the proposed interpretations must be associated with textual contexts that justify them. These interpretative choices can be discussed and justified if necessary.

The second type of validation is the one provided by experts in the domain. One of the characteristics of specialised corpus analysis is that the analyst in most cases is not a domain expert and is thus not familiar with the terms or phrasing used by experts. This means that data may be misinterpreted. It is therefore necessary to work closely with experts in the domain, who analyse a set of contexts that have been pre-selected either because of their difficulty or importance. Together, they can thus co-construct an interpretation that makes use of linguistic knowledge and knowledge of the domain. Validation by experts then often proceeds in a co-construction process (Picton 2009, 2014b).

A third type of validation can be performed by end-users, when the project call stems from an applied need. From a strictly lexical point of view, these users

are language professionals, such as translators, technical writers, or even computer scientists who need resources to make the tools work. Nevertheless, end-users can also be LSP teachers, non-experts wishing to access simplified definitions, experts from other disciplines, etc.

Therefore, the results are not only a set of terminology entries. They can also be data interpretations based on the study of terms. The way in which these results are presented may fall under what, in ergonomics, is called ‘usability’. This part of the validation can therefore be performed in association with ergonomists, which also situates Textual Terminology within the framework of Ergonomic Linguistics (see Section 4.1.2).

3.1.6 *Integrating the results into linguistic paradigms*

From a theoretical point of view, the consideration of extra-academic needs presents a risk. Even though the results might fulfil these needs in an *ad hoc* manner, their contribution to Linguistics might be overlooked. On the one hand, the results obtained might not be sufficient to enrich the description of terms and their use. In fact, it goes without saying that if results are not elaborated, they cannot be reused to deal with other similar problems. For instance, in the study of determinologisation, the results obtained are relevant to neology. In the case of popularisation, they pertain to shifts in meaning (e.g., Condamines, Humbert-Droz and Picton 2021; or Ledouble 2020). The integration (and theorisation) of these results in linguistic paradigms is thus of major importance.

3.2 Skills and protagonists

Echoing what has been described above, the following types of protagonist may be involved in this process (Table 1):

- Terminologists-linguists, who organize the work and conduct the analyses. From the perspective of a scientific study, they act as intermediaries between the extra-academic world and linguistic knowledge.
- Domain experts, who help to compile the corpus, validate the interpretation of the results and, in some cases, validate the results within the context of the initial project goals.
- End users, who may work in companies and are often the source of the demand, driven by quality departments or design offices.
- Terminologists-linguists may call upon other types of skill, e.g., those of NLP experts to adapt computer tools to the needs of the analysis.
- Ergonomists and their expertise can be used to characterize the need and evaluate the usability of the proposed results.

Since the early 1990s, these principles have remained relatively stable. However, several aspects need to be re-examined. On the one hand, a major evolution has taken place in NLP and Knowledge Engineering, thanks to the availability of huge amounts of data on the Internet. On the other, studies of specialised corpora have diversified. The following section describes the evolution of Textual Terminology and the perspectives that have emerged.

Table 1. Tasks/stages and protagonists – Textual Terminology

Tasks/stages	Protagonists
Identification of a need or a problem (translation into a linguistic issue)	(ergonomists) End-users Domain experts Language experts (analysts, terminologists-linguists)
Constitution of the corpus	Language experts Domain experts
Selection and use of tools	Language experts If necessary, NLP experts
Interpretation of the clues provided by tools	Language experts
Construction of a final interpretation	Language experts Domain experts
Integration of results into a linguistic paradigm	Language experts
Validation	(Ergonomists) Domain experts End-users

4. Textual Terminology today: Challenges and perspectives

Over the last thirty years, Textual Terminology has evolved considerably in regard to the following: (i) emergence of applied needs; (ii) development of Corpus Linguistics; (iii) cross-fertilisation with NLP, KE/AI. This section reappraises these three very closely linked areas, as they are today. This sheds light on the future of Textual Terminology and its development in the short and medium term.

4.1 New needs and perspectives

4.1.1 *New needs in Terminology*

Whereas research initially focused on the creation of termino-ontological resources (see Section 1), from the early 2000s onwards, the need to update these resources

began to surface (e.g., Condamines, Rebeyrolle and Soubeille 2004). This question led to the consideration of diachronic variation (see Section 2).

From then, new needs gradually emerged, which involved the study of other types of dialogical variation (Freixa 2006). As explained in Section 2.1.1, a project request from the CNES led to the integration of the analysis of diastatic variation. At the same time, another project was initiated by the Geneva Centre for Humanitarian Studies.² Its objective was to highlight differences in term usage from different perspectives in the field, in English³ (Egger, Picton and Schopper 2018). Diatopic variation is in this case important. We can also mention a study that we carried out on exobiology (funded by the CNES), an emerging discipline that involves different disciplines. This work shed light on the fact that terminological variation was not only a difficulty for interdisciplinarity, but that it could contribute to identifying the dynamics between multiple points of view regarding a concept (Condamines 2014).

In addition to envisaging other types of variation, this led us to reflect on the simultaneous analysis of different kinds of variations. In the case of determinologisation, for example, variation is both the diastatic (movement of terms from specialised texts to general texts) and the diachronic (movement of terms over time). This implies working on fairly complex comparable corpora, which are analysed, based on different types of variation.

In these analyses, the general methodological approach remains the same as the one described in Section 3. However, in addition to the overall difficulty of ensuring the representativeness of the data and their balance in the different sub-corpora (Section 3.1.1), the consideration of these more complex comparable corpora also highlights the issue of data availability. This question is directly related to the NLP tools and methods that Textual Terminology can use with these corpora (see Section 4.2).

4.1.2 *Towards new kinds of need: Controlled natural languages*

As a method of systematic linguistic analysis with NLP tools in specialised domains, Textual Terminology can be implemented in different ways. Its methodology can also be adapted to elements other than lexical units, such as the definition of Controlled Natural Languages (CNLs). CNLs are linguistic recommendations intended to limit risks related to the use of the natural language, particularly in

2. The URL for the Geneva Centre for Humanitarian Studies (formerly known as CERAH) is the following: <https://humanitarianstudies.ch/>.

3. For an interesting view on term usage: Do we all speak and practice “humanitarian” in the same way? (<https://evidenceaid.org/do-we-all-speak-and-practice-humanitarian-in-the-same-way/>). Accessed 21 February 2022.

companies and businesses (Ryan 2018, 294). Since this is also one of the functions of terminologies (Condamines 2010), many CNLs also have a terminology component. Like terminologies before the 90s, they are usually developed by experts in the domain, sometimes with the help of translators. This means that CNLs are mainly established by introspection, based on the competence of domain experts.

In another project with the CNES, we were asked to build a CNL for the writing of satellite specifications. We thus compiled a corpus of specifications written without instructions in order to identify their lexico-syntactic regularities (typical of that textual genre) and to use these regularities to construct the CNL.

We used the SDMC tool (Sequential Data Mining under Constraints) (Quiniou, Cellier and Legallois 2012, 167), on the corpus of specifications from the CNES. The tool was able to identify phrasemes, which not only include recurring lexico-syntactic structures but also multiword terms (for example, *en pointage inertiel* [inertial pointing]) and expressions. By deleting the terms present in the CNES database, we were able to maintain the structures that seemed to correspond to sentence structures (Warnier and Condamines 2015), for example [*sur réception de tc, le lvc [envoyer] ...*] (on receipt of the tc, the lvc [send]...). Based on the selected structures, we proposed alternative structures that conveyed the same information. In the example above, *envoyer* [send] may have the following forms: *envoie/doit envoyer/doit pouvoir envoyer/enverra* [send/must send/must be able to send/ will send]. Concerning injunctions, 20 sets of structures were thus specified in different lexical contexts. These structures were then proposed to experts in the writing of specifications, who evaluated their acceptability. These evaluations of spontaneous formulations (corpus) by means of acceptability tests (Warnier 2018) allowed us to propose usable recommendations. In fact, CNES decided to use our CNL to train the writers and editors of specifications.

These results point to the need for companies and businesses to employ linguists. The development of this CNL also led to a number of theses carried out at the CLLE laboratory and highlighted the need to borrow concepts such as 'usability' from ergonomics. The consideration of usability in the construction or adaptation of CNLs allowed Condamines (2021) to develop the concept of 'Ergonomic Linguistics'.

Ergonomic Linguistics provides a framework to propose answers to language needs, especially within firms. The aim is to propose linguistic standards (terminologies, CNLs, etc.) that will be effective in a given context as well as acceptable to users. From the perspective of Textual Terminology, (see Table 1), at the beginning of the study, Ergonomic Linguistics characterizes a need, and at the end of the analysis, it evaluates the relevance of the results in relation to this need and helps to make them available in a form that can be applied by the end-users. The integration of ergonomic skills and perhaps even cognitive psychology (to conduct usability testing) could be an interesting perspective for Textual Terminology.

4.2 Textual Terminology, NLP, and KE/AI: Towards a redefinition of the relationship

As previously mentioned in Section 1, corpus analysis for the compilation of termino-ontological resources was one of the founding pillars of Textual Terminology and of its connection with NLP and KE. However, the situation is quite different today. According to Condamines (2018c, 343), “the original ‘symbiotic relationship’ between Terminology and Knowledge Engineering, evoked by Skuce and Meyer (1991), no longer applies”. There are at least two possible reasons for this: (i) the loss of a common research objective; (ii) the question of corpus size and the methods used to explore it.

4.2.1 *From symbolic approaches to machine learning: Data quantity*

As pointed out by Condamines (2018c), research objectives have experienced a dramatic change. There is considerably less emphasis placed on building termino-ontological resources from texts, especially based on symbolic approaches, whereas more is placed on ontologies, based on automatic learning methods. The development of these methods has triggered a real paradigm shift in NLP. Even if some studies still use a symbolic approach to build ontologies (e.g., Asim et al. 2018), studies such as those by Church (2011) or Hall, Jurafsky and Manning (2008) have estimated that since the early 1990s, the use of statistical methods in NLP (including machine learning) has risen from 30% to over 90%.

The role of Linguistics in NLP has thus become less important. As Eensoo and Valette (2013, 2), point out, “[p]our une même application – par exemple la traduction automatique – les méthodes symboliques, jadis, mobilisaient pendant plusieurs années une armada de linguistes pour l’écriture de règles ; aujourd’hui, pour peu que des corpus parallèles de taille suffisante soient disponibles, un système par apprentissage nécessitera très peu de ressources humaines et de temps”.⁴ Efforts in NLP are now focused on developing these approaches, which most often require very large corpora.

Despite the lack of consensus on corpus size (e.g., Corpas Pastor and Seghiri Domínguez 2010), Fabre and Lenci (2015, 9) observe, “[t]here has been a clear shift from the treatment of middle-sized specialised corpora to the acquisition of distributional thesauri in the 90’s (...), to the compilation of corpora as large as possible, often heterogeneous in genre and domain”. In the context of (neural) machine

4. [f]or a single application – for example automatic translation – symbolic methods used to mobilize an armada of linguists for several years to write rules; today, provided that parallel corpora of sufficient size are available, a learning-based system will require very few human resources and time (our translation).

translation for instance, Ghaddar and Langlais (2020, 3595) explain that “[t]hese models require the use of large-scale parallel corpora to train millions of internal parameters”. In the field of automatic distributional analysis methods, underlying much current research in NLP (e.g., Heylen and Bertels 2016; Lenci 2018), Boleda (2020, 230) specifies that “[a] rule of thumb is to have at least 20–50 instances of each expression to represent; many languages, domains, or time periods simply lack these data.”

Whereas some authors seek to show that applications of current distributional approaches are possible in smaller corpora⁵ (e.g., Fabre et al. 2014; Tanguy, Brunet and Ferret 2020), these corpora still have a minimum of several million words, which is still quite large in comparison to the usual size of specialised corpora (e.g., L’Homme 2020a).

Today some (non-exhaustive) examples of large specialised corpora include the following: more than 23M words for the Ecolexicon corpus (León-Araúz, San Martín and Reimerink 2018); 85M for the *Humanitarian Encyclopedia*⁶ (Egger, Picton and Schopper 2018); 4.5M for the TALN corpus built by CLLE lab (Tanguy, Fabre and Bard 2020), etc. Moreover, with the development of new demands for terminology analysis (Section 3.1.), an increase in the size of specialised corpora is foreseeable. However, the following points about large specialised corpora should be highlighted:

- Specialised corpora are still relatively small in comparison to current approaches in machine learning (and especially distributional analysis in general language).
- Even if the total number of words is in the billions, the division into sub-corpora for the purposes of comparability considerably reduces the size of the datasets to be analysed. In the case of the Humanitarian corpus (85M words), the sub-corpora have between several millions and less than 1M words (Picton, Drouin and Humbert-Droz 2021).
- Specialised corpora are still difficult to collect and compile (Section 2.1.1).

Nevertheless, in view of new needs and the fact that data access is, to a certain extent, easier, Textual Terminology should tackle the issue of large specialised corpora, especially since it is closely linked to that of the NLP methods and approaches.

5. Special issues of periodic journals also question the place and assets of small corpora, for instance the CORPUS journal, issue 18 (<https://journals.openedition.org/corpus/3094>). Accessed 8 December 2021.

6. The Humanitarian Encyclopedia can be accessed at the following URL: <https://humanitarianencyclopedia.org/concept-analysis/>. Accessed 21 February 2022.

4.2.2 *Analysing specialised corpora today: Methodology*

The nature and evolution of specialised corpora raise different questions regarding the tools and NLP methods available for the analyses, and in particular, as they require the combination of quantitative analysis (e.g., to use data to identify leads and trends) and qualitative analysis (to refine the initial quantitative observations and interpret them appropriately).

This issue goes hand in hand with the need for data visualisation to analyse complex data (Picton, Drouin and Humbert-Droz 2021). Caple, Bednarek and Anthony (2018), for example, discuss this issue in relation to the analysis of data from different media (including text and images in news items). Anthony (2018) shows the diversity of possible means of visualisations and their impact on corpus analysis results. In *Diachronic Linguistics*, Hilpert (2011) proposes dynamic visualisation solutions to handle big diachronic comparable corpora for a fine-grained description of phenomena. Although these reflections seem to largely apply to the field of general language, Textual Terminology can greatly benefit from them.

Nonetheless, other approaches in NLP research seem far removed from the specific needs of Textual Terminology and specialised corpora. For example, distributional analysis approaches and techniques, which are at the heart of Textual Terminology methods (Section 3.1.4.), are developing exponentially in NLP, along with machine learning and big data possibilities. In Terminology, this path has been explored in some depth for semantic relation extraction (e.g., Bernier-Colborne and Drouin 2016, Bertels, this volume) and term extraction (e.g., Drouin, Morel and L'Homme 2020). However, regarding the overall corpus analysis, two issues arise:

- The complexity and diversity of distributional approaches make it often difficult for terminologists to understand how these approaches work and their applicability to the data at their disposal (e.g., Boleda 2020). Distinguishing the pros and cons of each approach requires advanced skills related to Computer Science, AI, and statistics.
- As previously mentioned, these methods are generally based on very large and heterogeneous general data, which do not reflect the context and needs of specialised corpora, even large specialised corpora. Current discussions in NLP do not therefore address certain issues that are directly related in Terminology.

However, various studies tend to show that the nature of corpora has an impact on the results provided by a machine learning approach in distributional analysis. For example, Del Tredici, Fernández and Boleda (2019) discusses recent methods for distributional analyses and short-term diachrony, whose relevance for Terminology is evident (e.g., Dury 2004, in this volume; Picton 2014b, 2018). These authors clearly highlight the impact of the temporal shift in corpora on the

results of diachronic distributional analysis and the difficulties of interpreting them. Choosing the right distributional method for the right corpus is thus of prime importance but remains a difficult point that requires navigating between the specific needs of Terminology and the evolution of NLP.

Regarding this distance between NLP and Textual Terminology, Firth's (1957) remark, "[y]ou shall know a word by the company it keeps", is very often cited to justify the distributional approach in NLP. This quote is often taken out of context, however, with respect to Firth's work as a whole. In another excerpt, quoted in Duvivier-Senis (2016), Firth describes the distributional approach as follows: "[f]irst the structure of **the appropriate contexts of situation** must be stated (our emphasis). Then the syntactical structure of the texts. The criteria of distribution and collocation should then be applied" (Firth 1968 [1952], 19). Various authors emphasise the same idea concerning the interpretation of the results provided by distributional approaches, such as in Fabre and Lenci (2015, 9):

What was clearly asserted in Harris' original method was the fact that such inductive semantic classifications reflected the Distributional Semantics Today's use of words in specific corpora. The approach was set in the context of the theory of sublanguages, based on the assumption that only **corpora from restricted domains** could guarantee the possibility to build up clear-cut semantic categories.

(our emphasis)

Consequently, there seems to be a certain distance between the original assertions by Firth and Harris (who had, however, different theoretical perspectives) on distributional analysis and current developments in machine learning and Artificial Intelligence on big data, which rely on large and heterogeneous data, by definition.

These elements fuel the reflection about the nature of the relationship between NLP and Textual Terminology today. While it is clear that the symbiotic relationship no longer exists, close collaboration seems possible and even essential. However, this collaboration must consider the specificities of specialised (large and complex) corpora and question the relevance and portability of technical developments in NLP for the new needs of Textual Terminology.

5. Concluding remarks

This chapter has proposed a retrospective look at the origins of Textual Terminology. After describing its methodological approach and the questions it raises, this perspective enabled us to highlight the evolution of this theory in the short and medium term, and where to situate it today.

The elements used to define Textual Terminology might initially seem quite similar to those of applied Corpus Linguistics (Hyland, Chau and Handford 2012) or even specialised discourse analysis (Gotti and Giannoni 2006). However, the specificity of Textual Terminology lies in the following: (i) it concerns only specialised corpora; (ii) the purpose of the study is considered throughout the analysis; (iii) terms are the point of entry into the texts; and (iv) tools are used as much as possible, with particular attention to the nature of the results obtained.

It also allowed us to highlight the cross-fertilisation between Textual Terminology and other theories of Terminology and, consequently, the ideas they share. We are referring to the need to base analysis on specialised corpora, the consideration of the situational context of project calls, the consideration of polysemy and variation, etc. (see chapter by Gaudin and Delavigne in this volume). Nevertheless, Textual Terminology, while nourishing these links, bases its specificity on four main axes: (i) the focus on tools for exploring corpora; (ii) the systematisation of its methodological approach (with a view to scientificity and reproducibility); (iii) the central role of experts (language professionals, users, domain experts, ergonomists) throughout the work process; and (iv) applicability of results to Linguistics.

This picture of Textual Terminology today, as well as of its needs and challenges, led us to highlight the new ways in which it can work together with other disciplines. For instance, within the framework of Textual Terminology, ergonomics and psychology have become of great importance to meet new needs (e.g., CNLs) and ensure the usability of the results obtained. Finally, the relationship of Textual Terminology with NLP and KE/AI has been transformed by the scientific goals and developments specific to each discipline. From being symbiotic, the modalities of this relationship have evolved into a real need for collaboration. Because of these expanded horizons, Textual Terminology is currently in rapid development and can look forward to a very bright future.

Terminology and Lexical Semantics

Marie-Claude L'Homme

Université de Montréal

The relationship between Terminology and Linguistics has fluctuated, depending on the prevalence of specific applications at a given time. A conceptual basis remains important in applications such as knowledge representation, and standardization, but other applications rely more heavily on corpora and must take into consideration the linguistic behavior of terms. This chapter lists questions that arise when analyzing terms in text and reviews some proposals made by Lexical Semantics to address them. Lexical semantic frameworks allow us to take into consideration terms that are often overlooked, such as verbs and adjectives. They also offer interesting ways to model various terminological relations (including collocations). Also emphasized is the fact that knowledge-based and lexicon-based approaches are only compatible to a certain extent.

Keywords: Terminology, Lexical Semantics, argument structure, lexical relation, verb, adjective, collocation, polysemy

1. Introduction

The relevance of the connection between Terminology and Linguistics has been assessed quite differently in the literature since Terminology was first recognized as a discipline. Some scholars take this connection for granted since terms are linguistic units, and as such, are subjected to phenomena that affect other linguistic units (polysemy, variation, semantic and syntactic constraints). This view implies that frameworks borrowed from Linguistics (and more specifically Lexical Semantics) offer useful principles and methods to capture and model the properties of terms. For others, terms are items that express knowledge (in other words, labels for concepts). Their nature is thus tied to the knowledge structures of special subject fields, and for this reason, they should be considered to be outside the realm of Linguistics.

This opposition can be explained, at least partly, by the applications for which terminology work is carried out. Although a conceptual basis remains of utmost importance in applications such as knowledge representation, standardization and

language planning, other applications that rely more heavily on corpora, such as specialized lexicography, terminography, information retrieval, and translation, must take into consideration the linguistic behavior of terms and their interactions with other lexical units. Knowledge-based and lexicon-based approaches are designed to address issues raised by different applications and, by doing so, focus on specific aspects of terms or look at similar aspects from a different angle.

This chapter reviews some contributions of lexical semantic approaches to the description of terms, often against the general background of knowledge-based approaches. Readers who wish a more in-depth description of knowledge-based approaches to Terminology should consult the chapters by Candel, Humbley and Meyer (this volume). Section 2 introduces an example that illustrates the kinds of questions that the analysis of a short extract can raise from a terminological viewpoint. The example then serves to showcase how lexico-semantic approaches address these questions (Section 3). Section 4 explains that, although both knowledge-driven and lexicon-driven approaches are necessary to understand the nature of the term, the paradigms on which they are based are only compatible to a certain extent and cannot always be combined.

2. Basic questions from the point of view of a specialized text

Example (1) is an extract of a Wikipedia entry on pollutants (all the links and footnotes were removed) that is used to illustrate the kinds of question that terminologists have to deal with when using texts to collect terms and information about them. The extract is assumed to be part of a larger corpus assembled for a dictionary compilation project. The four questions listed are those that usually arise when identifying terms in running text. Of course, some of these questions are also relevant for applications other than specialized dictionary compilation, such as general-language lexicography, translation, or even knowledge organization. However, the answers to these questions can vary – sometimes considerably – from one application to another.

- (1) A pollutant is a substance or energy introduced into the environment that has undesired effects, or adversely affects the usefulness of a resource. A pollutant may cause long- or short-term damage by changing the growth rate of plant or animal species, or by interfering with human amenities, comfort, health, or property values. Some pollutants are biodegradable and therefore will not persist in the environment in the long term. However, the degradation products of some pollutants are themselves polluting such as the products DDE and DDD produced from the degradation of DDT.

Pollutants, towards which the environment has low absorptive capacity, are called stock pollutants (e.g., persistent organic pollutants such as, non-biodegradable plastics and heavy metals). Stock pollutants accumulate in the environment over time. The damage they cause increases as more pollutant is emitted, and persists as the pollutant accumulates. Stock pollutants can create a burden for the future generations, bypassing on the damage that persists well after the benefits received from incurring that damage, have been forgotten.

Notable pollutants

Notable pollutants include the following groups

- Mercury Hg
- Persistent organic pollutants POPs
- Ozone Ozone
- Particulate matter PM
- Environmental Persistent Pharmaceutical Pollutants EPP
- Polycyclic aromatic hydrocarbons PAHs
- Volatile organic compounds VOCs

‘Pollutant’ in Wikipedia (2020)

Question 1: Which units or expressions correspond to domain-specific terms?

Answering question 1 is much easier for some units than for others. Among the easy cases, we can mention:

- *Pollutant* should be considered a term as it is the topic of the extract.
- *Polycyclic aromatic hydrocarbon* should also be defined as a term as it combines different properties that trigger the idea of “being specialized”: its structure is often associated with the idea of “being scientific” (i.e., neoclassical formation); it is linked to an acronym in the text (*PAH*); it is classified by the author as a type of pollutant (a unit which we already admitted was a term).

However, the status of other units is much more difficult to establish, based only on intuition. In some cases, it depends on the way the domain was defined prior to the selection. For instance, *product* and *substance* are used as generic terms for pollutants, but they might be considered too general if the domain was delimited as pollution.

In other cases, the difference with common usage may be felt to be too tenuous to validate a terminological status. *Accumulate* and *persist*, for instance, are events that are clearly linked to pollutants in the extract, but many might not see a clear difference between the meanings conveyed here and other more familiar meanings.

Interestingly, question 1 leads to other more specific questions that are listed below.

Question 2: Can verbs and adjectives be defined as terms?

Looking at the extract (1), it would be difficult to overlook the adjectives *biodegradable* and *non-biodegradable* since their meanings are clearly linked to pollution or more generally the environment. This being said, one might be tempted to consider that the adjectives are components of longer multi-word terms, e.g., *non-biodegradable plastic* (instead of *non-biodegradable* as an autonomous unit). Then, how should *biodegradable* be handled? If the longer sequence *biodegradable pollutant* qualifies as a term, the question is whether *pollutants are biodegradable* would be regarded as a variant.

Additionally, some terminologists might consider that the verbs *accumulate*, *introduce* and *persist* are a rather special case in this text and be tempted to regard them as terms (while other terminologists, as was pointed out above, might not perceive a real difference in meaning and decide otherwise). *Persist* is especially interesting since it is formally and semantically linked to *persistent* that also appears in the text and whose relationship with the domain is intuitively clearer. Similarly, the adjective *polluting* is closely linked to *pollutant* which was listed as a non-problematic term.

Assuming that verbs and adjectives can have terminological status raises the issue of accounting for them in specialized resources. Most verbs denote events and many adjectives denote properties and thus differ from nouns that denote entities. In fact, it should be pointed out that many nouns also denote events (*degradation* of X) or properties (*persistence* of X). Terms that denote events and properties are predicative and accurate descriptions should include a statement of their argument structure (X *persists* in Y; *persistent* X). Furthermore, it has been argued that many nouns denoting entities behave like predicative units, i.e., quasi-predicative units (Polguère 2012). *Pollutant* and *polluter* would belong to this category (X *is a pollutant* of Y; X *is a polluter* of Y). Describing predicative and quasi-predicative units requires an analysis of the linguistic contexts in which they are realized along with their arguments, as well as a close examination of their syntactic configurations (X *is persistent*; *persistent* X).

Question 3: When do multi-word sequences correspond to terms?

Part of the discussion for question 2 raises the issue of setting boundaries for multi-word sequences. Should *biodegradable*, *non-biodegradable*, *plastic* and *pollutant* be defined as single-word terms or should they be considered part of *non-biodegradable plastic* and *biodegradable pollutant*? The same question applies to *persistent organic pollutant*, especially as it is also compositional and *persistent* modifies another noun in the extract. In this case, however, the multi-word sequence is associated with an acronym (POP), and this could be evidence of the special status of this expression in the domain.

Question 4: How informative are lexical relations with regard to terminological status and meaning?

This question was addressed indirectly in the preceding discussions. For example, *polycyclic aromatic hydrocarbon* is classified by the author of the extract as a type of pollutant, which is a taxonomic relation. It would then be odd to define *pollutant* as a term, but not *polycyclic aromatic hydrocarbon*. Furthermore, the terminological status of *polluting* could be supported by its formal and semantic relation to *pollutant*. Although, the relationship is different – *polluting* could be characterized as “the typical property associated with *pollutant*” – it is logical to consider *polluting* to be a term as well. In addition, given the close relation between the verb, *persist*, and the adjective, *persistent*, they should be given the same status (terminological or non-terminological).

3. Lexical Semantics to address terminological problems

The four questions listed in Section 2 are by no means unusual or specific to this extract on pollutants. The notion of ‘term’ is far from consensual, and it has thus been defined in different ways. There are various reasons for this, some of which have already been mentioned (applications, knowledge-driven versus lexicon-driven perspectives). Not surprisingly, this has led to the emergence of various theoretical stances in Terminology (as evidenced by the contributions in this volume). The following subsections focus on Lexical Semantics approaches with a few references to knowledge-driven approaches when they can be used to highlight some differences.

3.1 Is the term a special kind of lexical unit?

In knowledge-driven approaches, the notion of ‘term’ depends on that of ‘concept’, which is defined, based on its position in a consensual knowledge structure. The term binds a concept to a linguistic label or designation. In such approaches, terms (or more precisely concepts, since designations and concepts can be considered separately, at least theoretically) are defined from within a special subject field and not as units of language per se. Hence, experts should be better placed than any other person to identify and define terms. However, this approach fails to explain how the term differs from other linguistic units, and more specifically ‘words’ or ‘lexical units’.

Early on, lexicologists, lexicographers, and language for special purposes (LSP) experts sought to characterize the difference between words (lexical units that pertain to general language) and terms (units that pertain to specialized domains)

(Guilbert 1973; Rey 1979, among others). This topic has also been addressed more recently (e.g., Petit 2001; Béjoint and Thoiron 2010), showing that it is still a subject of debate. This section presents three viewpoints on the matter: Terms are linguistic units with a special function (3.1.1); Terms are integral parts of the lexicon, and represent (formal or semantic) additions to the lexical stock of a language (3.1.2); Terms are lexical units conceptualized from the perspective of a specialized situation (3.1.3).

3.1.1 *Terms are linguistic units with a special function*

Early work (Guilbert 1973; Rey 1979) – a view that can still be found in more recent literature – considers that the special nature of terms lies in their function, i.e., the communicative purpose for which they are used. This view is the most compatible with knowledge-based approaches to Terminology. Accordingly, the term differs from the ‘word’ because it refers to a specialized entity or fact. It is thus closely tied to the real-world reality that it designates whereas a word conveys one or several meanings. Guilbert (1973, 10) states:

... le terme technique se définit par rapport à l’usage que l’on fait de la chose, aux composants de l’objet, aux caractères perçus par les sens (forme, couleur, dimension), à la localisation géographique, ou par référence à une taxinomie des êtres de la nature, dont l’ensemble constitue un champ sémantique ...¹

This function would explain why terms embody certain specific linguistic properties, such as the following: (i) linguistic realization in the form of nouns; (ii) tendency to be monosemous; (iii) prevalence of certain structures in scientific and technical domains (e.g., use of acronyms, neoclassical formation, multi-word sequences, symbols); (iv) possible internationalization; (v) language-independence of many concepts; (vi) potential standardization, etc. (Guilbert 1973; Muller 1985; Sager 1990; Kocourek 1991). It was also pointed out that these are tendencies rather than properties as more information is needed to characterize the terminological content of a specialized text. Even though these features apply to some of the terms in Example (1) such as *PCB*, *polycyclic*, *volatile organic compound*, they are not applicable to other units. For instance, *pollutant* and *persistent* sound familiar; *environment* is polysemous, and so on. According to Sager (1998, 41), “[...] the difficulty in distinguishing between two types of units stems from the fact that formally terms are indistinguishable from words”.

1. ... the technical term is defined according to the use that is made of the object, the components of this object, its characteristics that are perceived by the senses (shape, color, size), its geographical situation, or with respect to a taxonomy of living things, the whole of which forms a semantic field... (my translation).

Terms are integral parts of the lexicon

A second approach views terms as integral parts of the lexicon of a language and aims to understand their relationship with other lexical units. Rivenc (1971, cited in Muller 1985) represented the lexicon as a sun with rays and layers that overlap to a certain extent (Figure 1) (a similar representation appears in Rondeau [1984]). In Figure 1, domain-specific units are those that would appear in 1, 2, and 3, moving away from the center as they become more specialized. Components 1, 2 and 3 could each represent a different subject field since these fields share some lexical units in their less specialized areas.

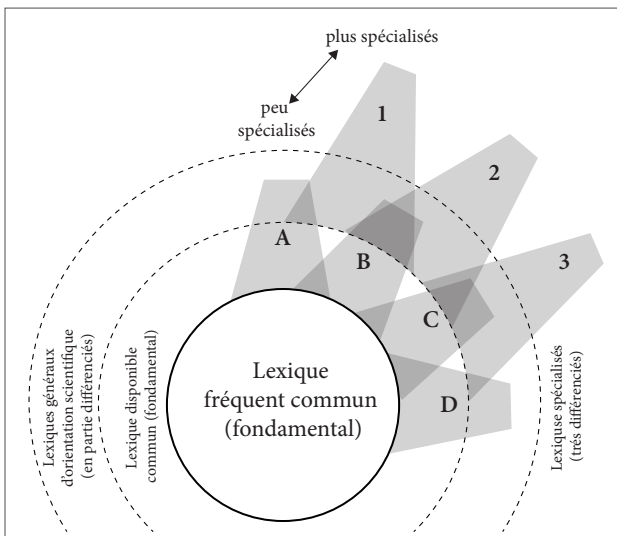


Figure 1. A sun representation of the lexicon (Rivenc 1971, cited in Muller 1985)

Of the approaches that consider the ‘term’ to be an integral part of the lexicon, a slightly different one asserts more strongly that terms are lexical units and defines terminological status from the perspective of polysemy (Kocourek 1991; Adelstein and Cabré 2002; L’Homme and Polguère 2008). This approach is reminiscent of a lexicographical practice in which usage labels are used to indicate domain-specific meanings. Kocourek (1991, 184) for instance, states:

... les termes ne représentent, potentiellement, que certaines acceptions de l’aire sémantique de l’unité lexicale, à savoir celles qui sont définies par les spécialistes dans les textes spécialisés.²

2. Terms can represent part of the senses (acceptations) in the semantic potential of lexical units, i.e., senses that were defined by experts in specialized texts (my translation). It should be pointed out that Kocourek uses *lexical units* to refer to polysemous lexemes.

Adelstein and Cabré (2002) formulate this somewhat differently since they state that terms are lexical units that have a terminological potential that can be activated in specialized communicative situations.

In their reflection on methods to combine lexicographical and terminographical resources, L'Homme and Polguère (2008) assert that the lexicon is a structured entity and that terms and lexical units can be described with similar descriptive tools. The authors identified different kinds of relationships that can be shared by specialized and general lexical units within the lexicon. What follows is a simplified account of the cases identified and the examples were adapted.

- ‘Pure’ terms that only appear in a specialized domain (e.g., *polycyclic*).
- Addition of new meanings to existing lexical units (specialized or not). This applies to *environment* whose basic meaning is “conditions, circumstances surrounding something of someone”, but is defined in computing as “a computer interface” and in the environment as “a global set of meteorological, biological conditions ...”³
- Specialized lexical units that can be used as such or slightly altered in general language. *Ozone*, *pollutant*, and *biodegradable* are now common in general usage. This migration from specialized to general usage can lead to a phenomenon called *determinologization* (Meyer and Mackintosh 2000), in which units, originally well-defined in a domain, can migrate to general usage and have their original meaning altered in the process.
- Specializations of existing meanings. In some cases, although there is no clear evidence of polysemy, there seems to be something special about the use of a lexical unit in a specialized text. For example, even though *accumulate* and *persist* express events that apply to pollutants, their meaning does not substantially differ from meanings with which we are familiar.

3.1.2 *Terms are lexical units conceptualized from the point of view of a specialized situation*

The last category identified by L'Homme and Polguère (2008), i.e., specializations of existing meanings, concern many lexical units in specialized corpora and could lead to the problematic conclusion that all lexical units can lend themselves to a specialized reading (a view that is held in some corpus-based approaches).

3. Polysemy also occurs quite often within specialized domains and manifests itself in different forms (L'Homme 2020b). Cases of regular polysemy (Apresjan 1974) are common, such as activity-result polysemy *pollution*₁ (*these include forest fires, floods, oil spills and pollution of waterways*); *pollution*₂ (*extensive inshore and coastal pollution*). Furthermore, alternations resulting in meaning distinctions are also quite common, such as the agent-instrument alternation: *pollute*₁ (*Pesticides pollute waterways and can harm animals and other plants*); *pollute*₂ (*We are destroying the earth by polluting the atmosphere with toxic emissions*).

Cruse (1995) provides a relational explanation that sets some helpful boundaries. He introduces ‘microsenses’, which, in contrast to standard polysemy, are not completely incompatible since they can be linked to the same superordinate. However, microsenses remain mutually exclusive since they can hold paradigmatic relations with different sets of lexical units. With the example of *knife*, Croft and Cruse (2004, 127) explain that although *knife* usually denotes a kind of instrument (= not completely incompatible), it can be first linked to *cutlery*, *fork*, *spoon* and, secondly, to *weapon*, *gun*, etc. (among other readings) (= mutually exclusive).

The microsense explanation seems to apply to the verb *persist* mentioned earlier. For instance, Examples (2) and (3) show very subtle differences between the environmental reading of *persist* and another reading.

- (2) *Sadly, these rumours are likely to PERSIST for a couple of weeks.*
- (3) *... these pollutants will not PERSIST in the environment in the long term.*

Although in the second context, *persist* expresses an event linked to pollutants, it is arguable that its meaning differs sharply from the *persist* that applies to *rumours*. In Examples (2) and (3), *persist* conveys the general meaning of “continue to exist”. However, *persist* in Example (2) belongs to a paradigm in which *last*, *go on*, *continue*, and *stop* would also appear. *Last*, *go on*, and *continue* could be defined as near synonyms and *stop* as an antonym. In contrast, *persist* in Example (3) belongs to a different paradigm since it concerns an event that applies to specific entities. In this regard, *accumulate*, *absorb* and *introduce* are more compatible with this usage of *persist* since they express events that concern the presence of pollutants in an area, albeit from different perspectives. *Accumulate* conveys the idea that the amount of substance is growing; *absorb* places the area in the foreground and the pollutant in the background (a sort of converse relation); finally, *introduce* involves an additional argument, i.e., an agent that places the pollutant in the area.

The discussion about *persist* naturally takes us to another explanation about the opposition between general and specialized usage, based on Frame Semantics (Fillmore 1982; Fillmore and Baker 2010). In a way, Frame Semantics attempts to reconcile the previous approaches since it establishes a link between function and lexical units. Furthermore, it allows us to account for specialized usages that cannot be defined solely on the basis of standard polysemy.

Frame Semantics is a cognitive framework that assumes that the meanings of lexical units are construed against the background of experience, beliefs or practices. While this view does not only apply to terms (or even lexical units for that matter), Fillmore (1982, 127) comments on an example in which he argues quite convincingly that differences can be observed between common and specialized situations:

In both everyday language and legal language there is a contradictory opposition between INNOCENT and GUILTY. In everyday language, the difference depends on whether the individual in question did or did not commit the crime in question. In legal language, by contrast, the difference depends on whether the individual in question has or has not been declared guilty by the court as a result of a legal action within the criminal system.

This same reasoning can be applied to *persist* in Examples (2) and (3). *Persist* in context in Example (2) involves a duration and location. However, this location is not explicitly mentioned and can be conceptualized as something abstract. *Persist* (Example (2)) can also be associated with a negative connotation. *Persist* in Example (3) also involves a duration and location: however, in this case, the location is concrete and appears to be more salient than the duration. These differences in conceptualization also result in *persist* holding different sets of relations, as previously mentioned. In terms of Frame Semantics, *persist* (Example (2)) and *persist* (Example (3)) would be modelled in separate frames that would account for these different situations. Concretely, in FrameNet (2020), *persist* (Example (2)) appears in a Process_continue frame (along with verbs such as *continue* and *go on*), which is linked to other frames such as Process_start, Process_resume, Process_end. It would be extremely odd to link a frame accounting for a *persist* (Example (3)) situation to these other frames.

3.2 Can verbs and adjectives be terms?

Even though Section 3.1 gives examples of verbs and adjectives,⁴ it did not address their possible terminological status. Again, the status given to linguistic expressions depends on the terminological approach taken. In those approaches that view terms as units with a designation function, an assumption shared by knowledge-based approaches, terms are strongly if not exclusively associated with the part of speech of nouns (single-word nouns or noun phrases). They can also be abbreviations or symbols that behave as nouns.

Guilbert (1973), Rey (1979) and Sager (1990) and other authors all acknowledge the special status of certain verbs and adjectives, but only do so if a close link can be identified with noun terms (for instance, when verbs and adjectives are derived from noun terms). Hence, *environmental* could be considered to be a term, since it is morphologically and semantically related to *environment*.

4. Adverbs are not mentioned in this section as they have not been dealt with specifically in Terminology literature. However, it can be surmised from the discussion that some adverbs can also have a terminological status: *-ly* adverbs derived from adjectives (*global* – *globally* in the environment) or adverbs that are formally identical to adjectives (*online* adjective in *online resource*; *online* adverb in *go online*).

That said, intuition tells us that some verbs and adjectives seem special when regarded from the perspective of a specialized text. Examples such as *biodegradable*, *persistent*, *persist*, *accumulate* were mentioned in Section 2. While *biodegradable* is obviously linked to the field of pollution, it might be more difficult to say the same for *persist* and *accumulate*. The dependence on a noun term mentioned by Guilbert (1973), Rey (1979) and Sager (1990) also raises problems. More specifically, even though *persistence* is linked to *persistent*, it would be difficult to argue that the adjective is derived from the noun and not the other way around (and this observation does not even take into consideration the verb *persist*). The same applies to *polluting*: it would be odd to consider this adjective with respect to the nouns *pollutant* and *pollution* without first considering the verb *pollute* and the long list of units derived from it.

To account for verbs and adjectives in specialized resources, another common approach consists in viewing them as components of the larger phenomenon of phraseology. Verbs and adjectives are taken into consideration as collocates of noun terms (Pecman 2012a; Dechamps 2013; Buendía Castro, Montero Martinez and Faber 2014, among others). Hence, *persistent* and *persist* could be listed as collocates of *pollutant* (*persistent* ~; ~ *persists*). This avoids having to define their terminological status. This approach would entail handling *biodegradable*, *non-biodegradable* (and also *persist* and *persistent* for that matter) as collocates of various noun terms (*non-biodegradable plastic*, *non-biodegradable pollutant*, *non-biodegradable substance*, etc.). Finally, this raises the issue of having to distinguish collocations from multi-word terms. In other words, should we define *biodegradable pollutant* as a multi-word term or *biodegradable* as a collocate of *pollutant*?

This second approach allows terminologists to take into consideration a wider set of verbs than the first one. However, in the same way as the first approach, it considers verbs and adjectives only to the extent that they can be linked to noun terms (syntagmatically rather than paradigmatically). Verbs and adjectives remain 'second-class' units with regards to nouns terms and not given full terminological status.

Other authors have set out to understand the contribution of verbs to specialized knowledge (Condamines 1993; Lerat 2002; Lorente 2007) and have proposed different classifications. For example, using a financial corpus, Condamines (1993) identifies the following verb types: (i) verbs that are specific to finance (e.g., *débiter* – *to debit*); and (ii) verbs that acquire a new meaning in this domain (e.g., *retirer* – *to withdraw*). In the domain of law, Lerat (2002) identifies three groups of verbs. Although his first and third groups are similar to those identified by Condamines (1993), he adds support or light verbs (e.g., *porter* in *porter atteinte* – *to undermine*, *to discredit*, lit. *to bring an attack*). Lorente (2007) establishes four categories: (i) true terms that convey domain-specific meanings; (ii) phraseological verbs that combine with a noun and both taken together can convey a specialized

meaning; (iii) discourse verbs used to organize ideas or argumentation; (iv) connective verbs that link concepts and provide information about the relations between these concepts.

An alternative approach argues that units belonging to different parts of speech can have terminological status. Again, the initial proposals in that direction focused on verbs (Picht 1987; L'Homme 1998, 2003; Lerat 2002). L'Homme (1998) lists a number of reasons why some verbs can be considered terms: verbs that are specific to the subject field; some noun terms derived from verbs and not the other way around; meanings that differ from common meanings; special status in a given domain, etc. English examples include *download*, *abort*, and *develop* in the field of Computing, and, *fragment*, *pollute*, and *persist* in the environment. In some domains, such as computing where activity concepts are prevalent, it would be odd to overlook verbs. It was even suggested that verbs can be an excellent starting point to discover the terminology of a domain. Durán-Muñoz and L'Homme (2020) showed that motion verbs are especially productive in the field of adventure tourism (*to trek*, *to hike*, *to climb*, etc.) and that these verbs lead to the creation of other terms (*hiking*, *hiker*, *climber*, etc.).

Regardless of the approach taken, there is a growing consensus on the importance of verbs in Terminology. An increasing number of verbs are recorded in specialized dictionaries and term banks. Similarly, the contribution of adjectives in conveying specialized knowledge is being dealt with in an increasing number of studies (Maniez 2015; Pitkänen-Heikkilä 2015; Bonnadona 2020). Moreover, much of what has been said about verbs also applies to adjectives in regard to their possible terminological status. However, adjectives raise two additional difficulties. Firstly, they do not form a homogenous set as some adjectives express attributes (*persistent* in *persistent pollutant*), whereas others express arguments (*pharmaceutical* in *pharmaceutical pollutant*). Secondly, since in knowledge-based approaches, noun phrases can be terms, the distinction between multi-word terms and collocations becomes a thorny issue.

The 'revised' status of verbs and adjectives has a number of consequences for terminological analysis. If verbs and adjectives can be terms, the criteria typically used to identify terms (establishing a relationship with a subject field) are much more difficult to apply. This is why authors have resorted to other methods to formulate new criteria. Such methods involve the analysis of lexical units within linguistic contexts (especially contexts that provide clues that help to define arguments structures of predicative units) or the use of paradigmatic relations. Three criteria, originally suggested for verbs (L'Homme 1998), can be applied to other lexical units.

Based on a short list of units whose terminological status is established within a domain (easy cases, such as *pollutant*), three relational criteria help define other

units that would otherwise be problematic. Using the extract in (1), let us assume that *pollutant* is a term, the following relational criteria can be used to identify other potential terms.

- A. **Nature of arguments.** Given a predicative lexical unit, such as *persist* that has two arguments (X *persists in* Y), if its arguments are realized in the form of terms (identified using other criteria), the predicative lexical unit probably corresponds to a term. Accordingly, the verb *persist* could be regarded as a term if its two arguments generally take the form of terms in a pollution corpus (see Examples (4), (5) and (6)).
 - (4) *Some pollutants are biodegradable and will not PERSIST in the environment in the long term.*
 - (5) *... toxins will PERSIST in water ...*
 - (6) *... E. coli can PERSIST in secondary, nonhost habitats ...*
- B. **Morphological relations.** If the lexical unit is morphologically and semantically related to a lexical unit that has already been identified as a term, based on other criteria, then it is probably a term as well. Accordingly, if *pollutant* is classified as a term, *polluting* should also be regarded as one since it is morphologically and semantically related to *pollutant*. The same criterion would also make *pollute*, *pollution*, *polluter* terms. Similarly, if *persistent* is regarded as a term, then *persist* would also be one.
- C. **Paradigmatic relations.** If the lexical unit is paradigmatically related to another lexical unit that has already been recognized as a term, based on other criteria, then the first lexical unit is probably a term. The criterion concerns paradigmatic relations that are not covered by criterion B, such as synonymy or near synonymy, antonymy, meronymy, and to a certain extent hypernymy. Assuming *pollutant* is a term belonging to the domain of pollution, then, *contaminant* (a near synonym) would also be a term. The same would be true for hyponyms of *pollutant* (*pesticide, fertilizer, chemical, carbon monoxide, etc.*).

Figure 2 shows how these lexico-semantic criteria were applied to identify and collect terms with *pollutant* as a starting point. In addition to new criteria for term identification, the assignment of a terminological status to verbs and adjectives raises the question of describing them in a way that is compatible with their linguistic properties. The majority of verbs and most adjectives are predicative units (as well as many nouns) and, as such, they have arguments. For instance, a description of the adjective *polluting* should account for the fact that it has one argument (*polluting* X (such as *vehicle, plant* or *substance*)). Similarly, should a resource include the verb *persist*, its description would account for two arguments: X (such as *pollutant, substance*) *persists in* Y (such as *environment, habitat, and soil*). Predicative

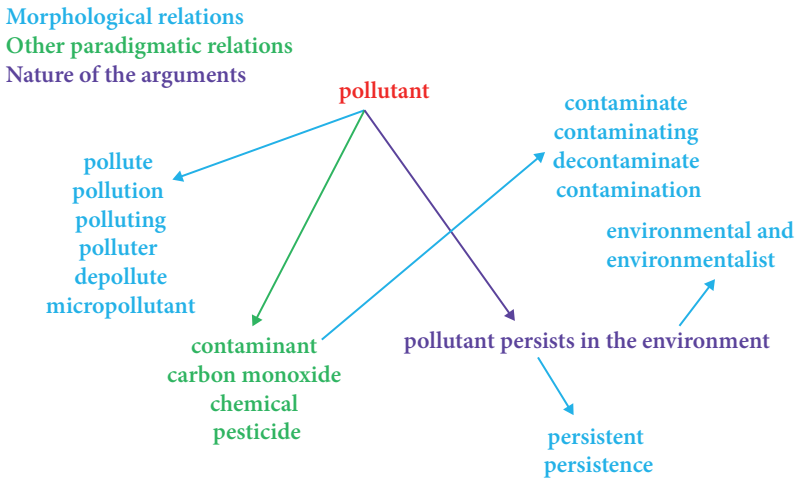


Figure 2. Application of lexico-semantic criteria to the identification of terms

units differ from nouns that denote entities, whose description is based on our conceptualization of a real-world object.

In line with this approach, various authors have suggested different ways of representing the argument structure of terms.

- Case Grammar: Condamines (1993) for finance (based on Fillmore 1968).
- Variables and/or typical terms: Binon et al. (2000) for business and L'Homme (1998) for computing (based on Melčuk, Clas and Polguère 1995 for variables).
- Semantic roles: L'Homme (2012) for environment terms (based on VerbNet 2020).
- Object classes: Lerat (2002) for law terms (based on Gross 1994)
- Frame Elements: Schmidt (2009) for soccer terms (based on Fillmore 1982; FrameNet 2020).
- Corpus Pattern Analysis: Alonso Campo and Renau Araque (2013) for environment terms (based on Hanks 2004).

A complete description of argument structures for predicative units is an interesting alternative to phraseological approaches. Instead of listing verbal or adjectival collocates in noun entries and having to replicate them (*biodegradable* under *pollutant*, *substance*, *plastic*, *matter*, *product*, etc.), the entry for *biodegradable* can state that it has one argument, and that this argument can be realized by a term that denotes a substance or material (*biodegradable* X (substance; matter)). This strategy is especially helpful for verbs since most verbs have more than one argument (X *persist* in Y). Phraseological perspectives listing collocates in noun entries (*a pollutant persists*) frequently omit important structural and semantic information.

3.3 Defining boundaries for multi-word sequences

Considering terms as lexical units generates an additional problem of having to define the boundaries of multi-word expressions that can be identified as terms. Knowledge-based and lexicon-based approaches differ quite sharply on this matter.

In knowledge-based approaches, multi-word units can correspond to terms as long as they label concepts in a knowledge structure. Thus we could define *biodegradable pollutant* and *non-biodegradable pollutant* as terms since these two labels identify specific categories of pollutants under which more specific substances can be classified. This is the option implemented in the GEMET Thesaurus (2020) where the concept labeled *biodegradable pollutant* is listed as a type of pollutant along with other types of pollutants (GEMET 2020) (Figure 3).

From a lexico-semantic perspective, this treatment of types of pollutants in the GEMET Thesaurus is problematic. The multi-word terms listed in Figure 3 are compositional and would pass all the linguistic tests designed to establish whether a sequence is compositional. We apply some of these tests to two compositional expressions (*biodegradable pollutant*; *air pollution*) and a non-compositional expression (*greenhouse effect*).

Definition

Any substance, usually a residue of human activity, which has an undesirable effect upon the environment.

[SHOW SOURCE](#)

Related terms

Related: [emission](#)

Narrower:

[air pollutant](#) | [bioaccumulative pollutant](#) | [biodegradable pollutant](#) | [biological pollutant](#)
 | [chemical pollutant](#) | [effluent](#) | [food pollutant](#) | [micropollutant](#) | [noise pollutant](#)
 | [non-biodegradable pollutant](#) | [physical pollutant](#) | [pollutant behaviour](#)
 | [pollutant elimination](#) | [radioactive pollutant](#) | [soil pollutant](#) | [urban pollutant](#)
 | [water pollutant](#)

Themes: [pollution](#)

Group: [WASTES](#), [POLLUTANTS](#), [POLLUTION](#)

Figure 3. Pollutant in GEMET (2020)

- Is insertion possible? If it is, then the sequence is probably compositional.
biodegradable, non-bio-degradable organic and non-organic pollutant
air, soil, and water pollution
**greenhouse or cooling effect*
- Is the contribution the same in other sequences? (Cruse 1986). If it is, then the multi-word sequence is compositional.
biodegradable pollutant, biodegradable substance, biodegradable matter
air pollution, water pollution, soil pollution
*greenhouse effect, *greenhouse impact, *warming effect*
- Is the contrast recurrent? (Cruse 1986). If it is, then the multi-word sequence is compositional.

These pollutants are biodegradable. These pollutants are persistent.

The biodegradable pollutants cannot be removed. The persistent pollutants cannot be removed.

Air pollution in this area is a serious cause of concern. Water pollution in this area is a serious cause of concern.

This country took strong measures to reduce pollution in the air. The country took strong measures to reduce contamination in the air.

For *greenhouse effect*, the only way to see a recurrent contrast is to replace the entire expression.

Accumulation of gases in the atmosphere enhances the greenhouse effect.

Accumulation of gases in the atmosphere enhances global warming.

a major environmental problem like the greenhouse effect ...

a major environmental problem like global warming ...

There is no simple solution to this problem as many compositional multi-word sequences would spontaneously be identified as terms by experts and non-experts alike: e.g., *persistent organic pollutant, volatile organic compound*. There seems to be a conflict between the lexical units that have been defined based on linguistic criteria and terminological units that are defined based on more intuitive perceptions of their 'specialization'.

Needless to say, the definition of terminologically-relevant multi-word sequences has important consequences on the way they are recorded in specialized resources. A knowledge-based resource will tend to list several entirely compositional sequences as terms as in the GEMET Thesaurus (see Figure 3). A lexicon-focused resource should record single-word terms and consider longer sequences if they are non-compositional (e.g., *greenhouse effect*). Other combinations can be taken into consideration, though they would be described differently. For instance, *soil*, in *soil pollutant* would be regarded as a possible realization of an argument of *pollutant* (*X is a pollutant of Y* (e.g., *food, water, soil*)) (see Figure 6). *Pollutant*,

on the other hand, would be defined as a possible argument of *biodegradable* (*biodegradable X*).⁵

3.4 Lexical relations and how they contribute to terminological analysis

Lexical relations, which have been mentioned throughout this chapter, can be used to validate terminological status or semantic distinctions, or to verify the compositionality of multi-word expressions. They include paradigmatic relations (taxonomies, synonymy or near synonymy, antonymy) and syntagmatic relations (predicate–argument relations and collocations). If there is one area on which knowledge-based and lexicon-based approaches agree, it is on the importance of ‘structure’ in Terminology. It is usually assumed that knowledge in domains is structured – by concepts in knowledge-based approaches or by terms in lexicon-based approaches. Regardless of the approach, the characterization and labelling of relations are important aspects of terminology work.

Knowledge-driven approaches have traditionally focused on generic-specific and part-whole relations. Synonymy, which is a relation that holds between designations rather than concepts, is considered when the terms designate the same concept (conceptual or true synonymy). In running text, concepts can be expressed in many different ways that include, but is not limited to, synonyms. Term variation was introduced and has become an active research topic in knowledge-based and lexicon-based work (see the contribution by Freixa, this volume).

Apart from generic-specific and part-whole relations, many other conceptual relations can be taken into consideration and some have been investigated more thoroughly in recent years (cause-effect and entity-function, for instance) (see chapters by Faber and Marshman, this volume). Sager (1990, 35) lists around 20 different relations, some of which are more relevant in certain subject fields than others: <material-product> (‘plastic’ – ‘container’), <process-product> (‘discharge’ – ‘pollutant’), <process-instrument> (‘pollution’ – ‘pollutant’), <object-operation> (‘chlorine’ – ‘disinfection’), etc.

A knowledge-based approach applied to the concept ‘pollutant’ would result in a representation similar to the one in Figure 4. The data come from the GEMET Thesaurus (Figure 3) and the representation is based on the content of definitions in the thesaurus.

5. Interestingly, some work investigates the semantic relations between components of multi-word terms. In some analyses, the predicative-argument relation between the components is inferred (see chapter by Faber in this volume).

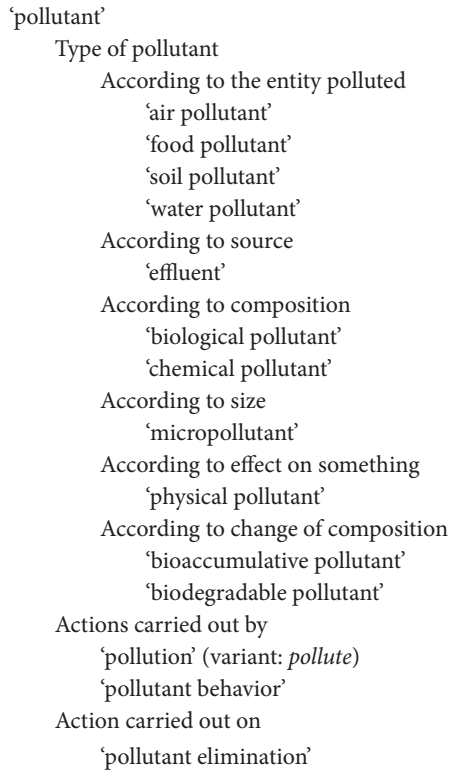


Figure 4. ‘Pollutant’ and related concepts

Even though structure and relations are equally important in lexicon-based approaches, the kinds of structure differ quite sharply from those discussed in this section so far. Accounting for all the relations in a domain can produce very complex networks (see Figures 5 to 7). The reasons for these differences were mentioned in previous sections.

- **Different parts of speech.** Verbs and adjectives appear in a wide variety of relations that differ from those in which nouns that denote entities appear. For instance, one of the meanings of *pollute* (as in *this substance pollutes the area*) is linked to:
 - *pollution* (defined as an activity): *pollute* and *pollution* carry the same meaning.
 - *pollution* (defined as an entity): *pollution* expresses is a result with respect to *pollute*.
 - *pollutant*: *pollutant* is the name of the substance that pollutes (the instrument or material).
 - *polluted*: this adjective applies to the second argument of *pollute*, i.e., the term that expresses the area.

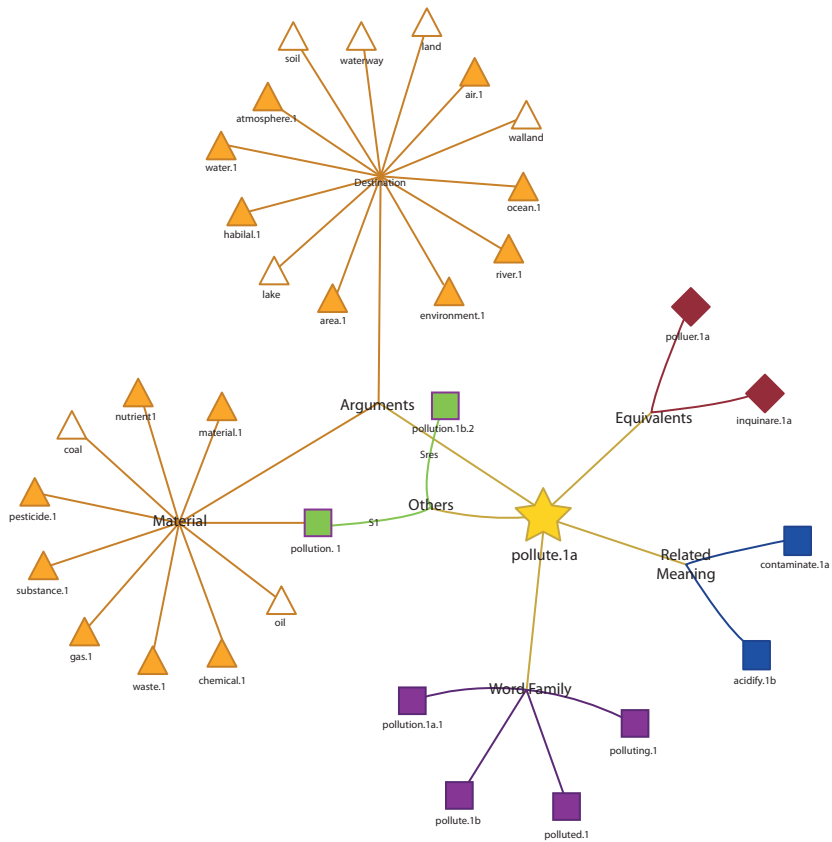


Figure 5. Relations held by *pollute*_{1a} with other terms (NeoVisual 2020)

- *polluting*: this adjective applies to the first argument of *pollute*, i.e., the term that expresses the substance. (see Figures 5 and 6 for relations of the two meanings of *pollute* with other terms).⁶

Considering verbs and adjectives in terminological analysis broadens the spectrum of relations in other ways, for instance:

- Adjectives and verbs often appear in oppositions, which are less frequent for nouns: complementary (*biodegradable* – *non-biodegradable*), gradable (*abundant* – *rare*), reversive (*deforest* – *reforest*), converse (*endangered*; *harmful*) (Gagné and L’Homme 2016).
- Near synonymy is more common between sets of adjectives or sets of verbs than true or conceptual synonymy between nouns that denote entities.

6. This first list points to argument relations (Melčuk, Clas and Polguère 1995).

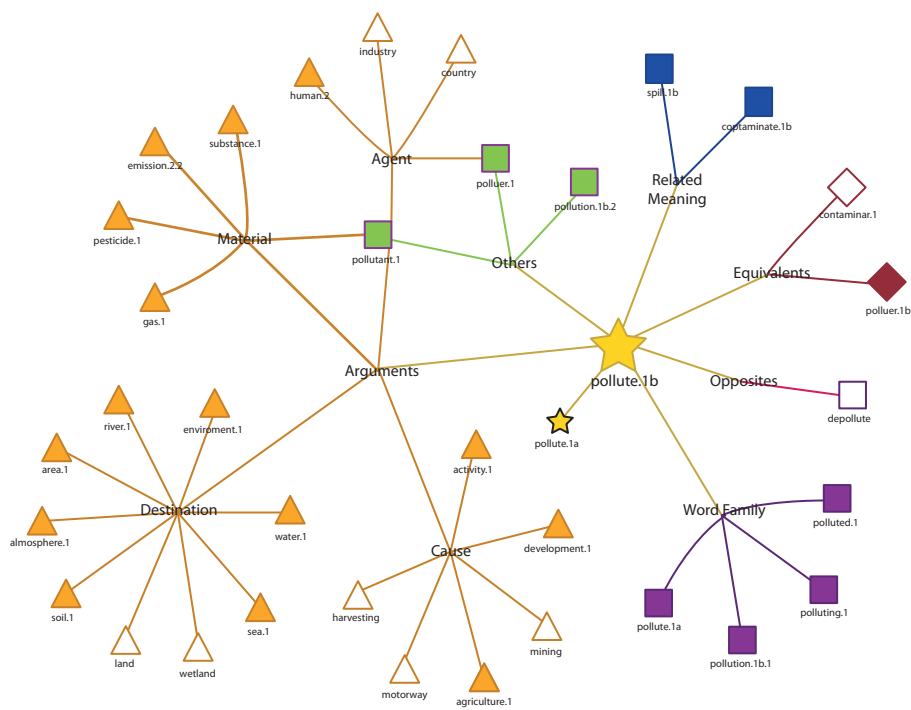


Figure 6. Relations held by *pollute*_{1b} with other terms (NeoVisual 2020)

- **Terms as lexical units.** As mentioned in the previous section, defining terms as lexical units impacts the way multi-word sequences are classified. While knowledge-based approaches consider many compositional sequences to be terms (e.g., *biodegradable pollutant*), lexicon-based approaches regard multi-word sequences as lexical units if they are non-compositional. This affects the description of relations in different ways. In knowledge-based approaches, *biodegradable pollutant* can be defined as a type of *pollutant* (a specific-generic or hypernymic relations as in the GEMET Thesaurus). A lexicon-based approach would attempt to define *pollutant* with respect to *biodegradable* (as a possible realization of its argument) or *biodegradable* would be defined with respect to *pollutant* (as a possible collocate).
- **Predicative units.** Predicative units and their arguments share different kinds of relations. We already mentioned *pollutant* with respect to *pollute* whose argument is instrument or material. In Figures 5 to 7, linguistic realizations of arguments are clustered under labels for semantic roles: Agent (e.g., *human*, *company*); Material (e.g., *chemical*, *gas*, *pesticide*); Destination (*atmosphere*, *environment*, *water*).

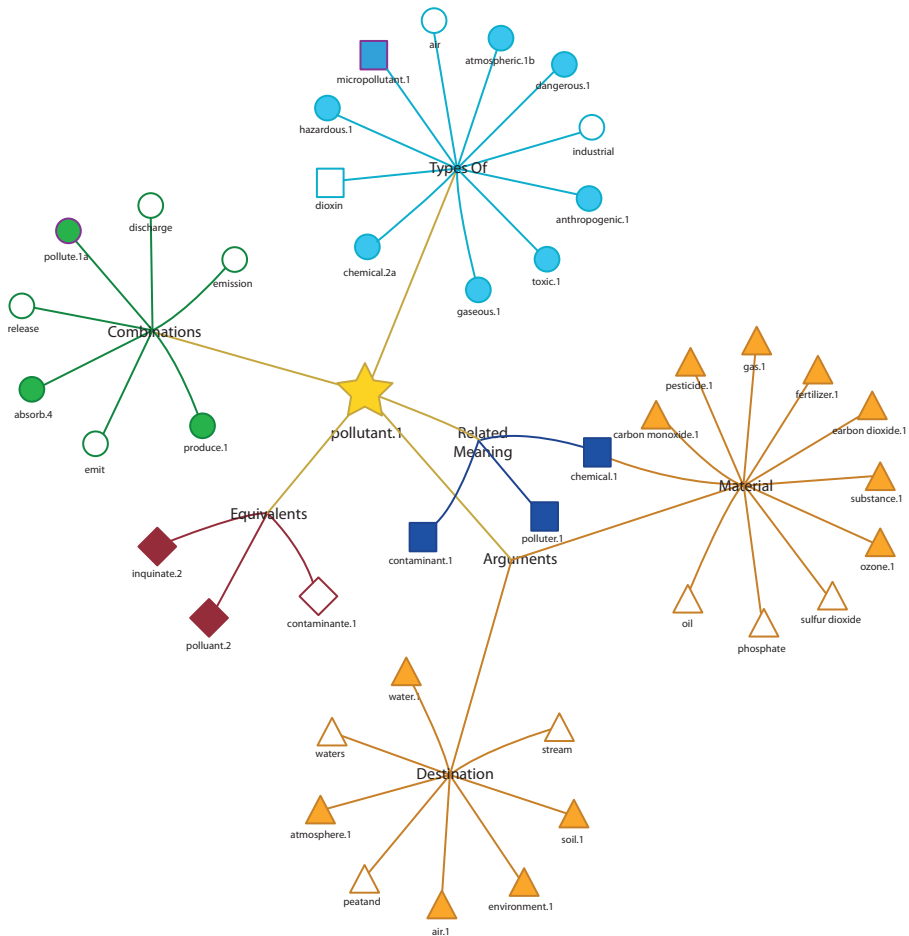


Figure 7. Relations held by *pollutant* with other terms (NeoVisual 2020)

- **Polysemy and relations between meanings.** Defining terms as lexical units and considering different parts of speech also has an impact on polysemy and how it is handled in terminological resources. In lexicon-based repositories, polysemy increases when compared to knowledge-focused resources. Furthermore, semantic distinctions are reflected in the relations that apply to one meaning with respect to another. Finally, relations can be established between the meanings of a polysemous items:
 - *pollution*₁ and *pollution*₂: activity-result.
 - *pollute*_{1a} and *pollute*_{1b}: An agent causes an instrument or material to *pollute*_{1a}.
 - *erode*₁ and *erode*₂: A cause causes an entity to *erode*₁.

The list of differences between knowledge-based and lexicon-based approaches to relations is much longer and it is impossible to exhaust it here (see L'Homme 2020a for a more detailed account). Instead, we will provide an illustrative example.

Figure 7 shows the relations that are encoded for *pollutant* in the DiCoEnviro (2020), a dictionary on the environment which is based on a lexicon-based approach. In contrast to Figure 4, in this representation, *water* and *soil* are described as arguments of *pollutant* (*pollutant of Y*), whereas *soil pollutant* and *water pollutant* appear as multi-word labels for concepts in GEMET. Certain relations, such as *chemical*, *contaminant* and *polluter*, are classified as paradigmatic (in squares), whereas others, such as *atmospheric ~*, *hazardous ~*, *absorb a ~*, *pollute ~ with ~*, are syntagmatic (in circles). Other differences include items that appear in one representation but not in the other. This concise comparison highlights the numerous differences in the description of relations when one approach is chosen instead of the other.

4. Concluding remarks

This chapter has presented different ways of addressing the notion of 'term' from the perspective of Lexical Semantics in comparison to knowledge-based approaches. When applied to terminological data, Lexical Semantics offers the means to account for different term types (nouns, verbs, adjectives, adverbs). It is also a valuable tool for the description of predicative terms and their argument structure, and can be used to model a vast array of paradigmatic and syntagmatic relations. Lexical semantic approaches are much better equipped than knowledge-focused ones to tackle questions linked to the behavior of terms in text. Approaches that exclusively focus on knowledge must consider terms as units that are emptied of their linguistic content in order to provide a consensual representation of the knowledge structure of a domain. Nevertheless, terms appear in textual environments and by ignoring their linguistic properties, much of what characterizes them is omitted.

Of course, no framework has all the answers since each focuses on the phenomena for which it was designed. It might be argued that lexicon-based perspectives are less grounded in knowledge and because of this, fail to provide a clear representation of the knowledge structure of a domain. Be that as it may, Lexical Semantics, and especially relational models (Cruse 1986; Melčuk, Clas and Polguère 1995), offer linguists and terminologists, who are typically not domain experts, different means to validate their intuitions (or "discipline" them, Cruse 1986). Among these, are different clues provided by the lexicon itself, i.e., relations that help validate terminological status or semantic distinctions. More recent

proposals can also help to capture and explain much subtler phenomena, such as microsenses. Lexical Semantics complement approaches that rely exclusively on expert knowledge, and can be used to mitigate the multiplicity of interpretations to which corpus-based approaches tend to lead.

Given the differences between the assumptions of lexicon-based and knowledge-based approaches and the principles on which they rely, the question is whether they can be used simultaneously in terminology work. In fact, an increasing amount of work combines different approaches in the analysis of terms (corpus-based analysis, use of Lexical Semantics, involvement of experts for validation steps, etc.). Typically, each project, depending on its objectives, eventually tips the balance slightly more towards knowledge or more towards the lexicon. At the present time, there is no way to capture the complex properties of terms and integrate all of them in a unified model. Nevertheless, there is an increasing awareness that Lexical Semantics is a very necessary part of the equation.

Acknowledgments

I would like to thank Pamela Faber and John Humbley for their comments on a previous version of this chapter.

PART 5

Corpus-based Terminology

Text genres and Terminology

Mojca Pecman and Natalie Kübler

Université de Paris

Terminology is intrinsically linked to textual genres. Text genres became an essential component of terminological research and practice in the context of Corpus-based and Textual Terminology. These approaches placed a spotlight on specific terminological features according to genre types and made contrastive analysis possible with emphasis on genres, registers and, consequently, domains. This chapter discusses the way Corpus-based and Textual Terminology have paved the way to placing genres in the core of terminological analysis. It explains why textual genres are often looked on as fuzzy categories. It also presents the various studies in Terminology involving textual genres and the general hypotheses, basic principles and the methodologies used therein. Finally, it examines the impact of text genres on theoretical and applied Terminology and on the future research in Terminology.

Keywords: text genres, corpus-based Terminology, Textual Terminology, textual corpora, genre-domain configuration, contextual analysis, theoretical and methodological principles, register

1. Introduction

Terminology is an intricate and interdisciplinary science (Humbley 2002). Its core object of study, terms, are the embodiment of knowledge and thus of crucial interest for specialised communication as well as for understanding and sharing information on any specific subject area. Terms also play an important role in the process of building and disseminating knowledge. It is thus not surprising that research in Terminology has branched out over the past decades to encompass the various approaches which, each in its own way, shed light onto the nature, behaviour, usage and utility of terms.

As research in Terminology has developed, it has become intrinsically linked to textual genre. In many respects, genres are crucial for language and discourse analysis, and particularly for corpus-based and LSP (Languages for Specific Purposes)

studies. Research in LSPs has developed genre analysis (cf. Swales 1990), and as terminology can be one of the defining features of LSPs, it is not surprising that it is also concerned with genres. This chapter thus focuses on the role of textual genres in Terminology with the objective of providing an overview of the basic principles and the methods used to integrate genres into terminological analysis and practice.

Text genres became an essential component of terminological research and practice in the context of the development of Corpus-based Terminology and Textual Terminology (see Condamines and Picton, this volume). Both Corpus-based Terminology and Textual Terminology are generally considered as a response to the shortcomings of the General Theory of Terminology (GTT). Back in the late 90's, the first works on the methodologies developed in Corpus Linguistics launched the exploration of what corpus-based terminography could offer the discipline for identifying terms and retrieving contextual fragments.

Genres are a dimension through which terminological data are prospected and analysed in textual corpora. They put the spotlight on specific terminological features according to various genres and make contrastive analysis possible with emphasis on text types, registers and, consequently, domains. As Swales (1990) indicates, LSPs belong to a specific discourse community that shares specific goals, but also specific communication features, such as a good command of certain textual genres, specialised vocabulary, and organisational criteria. So, each discourse community has its own specific genres or sets of genres. Thus, the choice of a genre or of a set of genres is correlated with the choice of a domain. However, textual genres are often perceived as fuzzy categories. Despite their importance and their complexity, there are almost no studies addressing the theoretical and methodological principles of working with text genres in the field of Terminology. One rare exception is Rogers (2000). The present contribution aims to fill this gap.

In Section 2, we discuss the way Corpus-based Terminology, in combination with Textual Terminology, has paved the way for considering textual genres as an important category for terminological analysis. The following section (Section 3) stresses the importance and the complexity of the category of genre. This is followed by a presentation of the typical approaches to text genres in various studies in Terminology and by the evaluation of the place they occupy therein. Section 4 is devoted to general hypotheses, basic principles and the methodologies used in these studies. Section 5 stresses the impact of text genres on theoretical and applied Terminology and on the future of research in Terminology.

2. Integrating text genres into Terminology research and practice

Text genres gained in importance in terminology research and practice in the context of textual and corpus-based approaches to Terminology. After the classical Wüsterian approach to Terminology (Wüster 1968) and subsequent socioterminological trend (Gaudin 1993), the textual and corpus-based approaches to Terminology began to be explored in the late 90s, respectively in works by terminologists such as Bourigault and Slodzian (1999) and Pearson (1998). The nineties was also the decade in which Corpus Linguistics became acknowledged as a theoretical and methodological approach, thanks to John Sinclair (1991), which allowed other researchers to apply Corpus Linguistics in other disciplines. Among them, we find Johns (1991) who started using corpora in what he termed *data-driven learning*, and Baker (1995) who was the first to use corpora in Translation Studies, and who initiated what is called today Corpus-Based Translation Studies (CBTS) research. CBTS was first applied to Descriptive Translation Studies as it considered translated texts as a research object. In applying Corpus Linguistics methods, Baker suggested that translated texts presented characteristics that differed from original texts, raising the hypothesis of translation universals. Furthermore, CBTS also started to explore corpus use in teaching specialised translation and, in this respect, have led researchers to study how parallel and, above all, comparable specialised corpora could be used to extract terms in the source and target language (among others at the beginning of the century, Faber 2002; Kübler [2003] 2014; Maia [2003] 2014). Automatic term extraction from corpora has also flourished as a research field with Daille (2001); Peñas, Verdejo and Gonzalo (2001); Drouin (2004), among others.

The textual approach to Terminology could be viewed as a response to the socioterminological approach (see Delavigne and Gaudin in this volume), which had been dominant until then and which emphasizes the circulation of terms in professional situations and various spheres of activities, thereby pushing spoken genres to the forefront and casting a shadow on textual varieties of specialised discourses. The principles of Textual Terminology stipulate that texts are to be considered as the basic material for retrieving and prospecting terms (Bourigault and Slodzian 1999; Slodzian 2000).

At the turn of the century, the essential role of texts in practical work and research on terminology was increasingly asserted (e.g., Bourigault and Charlet 2000; Aussenac-Gilles, Biébow and Szulman 2000a; Bachimont 2000; Jacques 2000). Moreover, in LSP communication studies, the role of specialized texts and specialized textual genres became one of the main focuses since Swales' (1990) study on discourse communities and genre analysis. According to Swales (1990, 9), "genres are the properties of discourse communities". He explains that "one of

the characteristics that established members of these discourse communities possess is familiarity with the particular genres that are used in the communicative furtherance of those [common] sets of goals”, genres being the “classes of communicative events that typically possess features of stability, name recognition and so on” (Swales 1990, 9).

Furthermore, the relation between text genres and terminology is bridged via the corpus. Thus, corpus-based and corpus-driven studies have enabled terminological analysis to be conducted in respect to specific text types. All the studies focusing on text genres integrate corpus-based methods. It is historically attested that the textual approach to Terminology is related to the development of Corpus Linguistics and computerized techniques for probing language units (cf. Kübler 2001). It is not surprising that both approaches, Textual Terminology and Corpus-based Terminology, appeared simultaneously.

Back in the late 90’s, the first works on terms in context (cf. Gerzymisch-Arbogast 1996; Pearson 1998) benefited from methodologies developed in Corpus Linguistics and launched the exploration of what corpus-based terminography can offer the discipline, at that time for the identification of terms and retrieval of contextual fragments. This work paved the way for Textual Terminology (see Condamines and Picton in this volume). They started pointing to the qualitative, semantic differences between terms represented in codified sources such as terminologies and specialized dictionaries, in comparison to their use in texts, “even in authoritative texts, which often weave their meaning in different ways by redefining key terms and concepts” (cf. Rogers 2000, 4). The terms in entries of an organized terminology and their embedding in running text are two different types of objects (cf. Rogers 2000).

The GTT and standardizing terminology work is thus opposed to the findings revealed by Textual and Corpus-based Terminology. For instance, standardizing terminology work looks to resolve the problems of synonymy and variation by decontextualizing terms (for instance *disease*, *disorder*, *illness*, *syndrome* (Rogers 2000)), and to establish equivalences across languages (e.g., GR: *Erkrankung*, *Krankheit*, *Morbus*, *Syndrom* ; FR: *maladie*, *désordre*, *syndrome* (Rogers 2000)) – a purpose also pursued in bilingual or multilingual lexicography – whereas the textual approach allows us to take into account terminological variation by placing it in a discourse structure and recording it in term bases with emphasis on contexts and genres which account for variation.

As Poibeau (2005) observes, at the beginning of Corpus Linguistics studies, the representativeness of the corpus was a question extensively debated (see also Nazarenko 2004). After more than three decades of research on and with corpora, the dominant consensus today consists in considering that a corpus can never be fully representative of a domain, but can only reflect a part of the language

phenomena under study (see Condamines 2003). This precision and this restriction are necessary for practical purposes. Likewise, certain notions or dimensions remain central and cannot be reduced or neglected when constructing the corpus, when searching the corpus and when making interpretations of the observations and results: these are genre, register and point of view (cf. Bouveret 1998).

Consequently, combining research in Terminology with studies on textual genres can be considered to be yet another response to the inadequacies of the General Theory of Terminology (GTT) for the future development of the discipline (cf. Humbley, this volume). As underlined by L'Homme (2004), the landmarking theoretical approach offered by the GTT, which was essentially standardization-driven, was partially adapted and partially abandoned because it would not offer suitable responses to many important questions that terms raise when investigated in texts and different situations.

Progressively, according to L'Homme (2006b, 181), “the situation has changed due to the number of factors, such as the amount of criticism the GTT has been subjected to (Bourigault and Slodzian 1999; Cabré 2003; Gaudin 2003; Temmerman 2000), proposals for incorporating Lexical Semantics in Terminology (Frawley 1998; Lerat 1988; Béjoint and Thoiron 2000; L'Homme 2003), work that integrate corpus-based methods (Ahmad and Rogers 2001; Meyer and Mackintosh 1996; Bowker and Pearson 2002) and computational approaches (Bourigault et Habert 1998)”. One of the best-known challenges for the GTT is Cabré's Communicative Theory of Terminology (CTT). Cabré (2003, 167) explains that the gap between the GTT and CTT lies in the fact that the former focuses on “what terminology should be in order to ensure unambiguous plurilingual communication”, while the latter is more “about what terminology actually is in its great variety and plurality”.

In this venture to characterise the role of textual genres to Terminology, one can notice that corpus-based and textual approaches seem less obviously characterised in the history of the theoretical evolution of Terminology. They are overwhelmingly documented in research in Linguistics, but they seem to be seriously overlooked in texts on theoretical approaches to Terminology. The theoreticians tend to distinguish the General Theory of Terminology, the Extended General Theory of Terminology, Socioterminology, the Communicative Theory of Terminology, Sociocognitive Terminology, Termontography, Ontoterminology, and Frame-Based Terminology (e.g., Cabré 2005; Sageder 2010). Whether corpus-based and textual approaches to Terminology fall into a category of the Communicative Theory of Terminology and/or Sociocognitive Terminology (cf. Temmerman, this volume) remains an open question, just as whether they can be considered as one rather than separate approaches. More plausibly, they could be considered to constitute a framework of specific tools and methods for prospecting terminology from any of the approaches listed above.

Indeed, in Terminology research and practice, the corpus-based approach, alongside the textual approach to Terminology, often coexist with other approaches. Their common theoretical and methodological bases make them good allies for all those who are interested in or in need of investigating terms in contexts. When the focus is on the text accessed through the lens of corpus – whether it be for writing, reading, or translating purposes – the question of genre arises quite spontaneously. Corpus-based Terminology alongside with Textual Terminology have offered a new route for analysing terms by allowing terminologists and linguists to survey them in relation to textual genesis and specific features of textual genres. Moreover, given the complexity of the notion of genre, terminological analysis applied to genre makes it possible to isolate and comprehend some crucial elements about the genres.

3. Genre: A fuzzy but essential category

Textual genres are generally defined as “conventional forms of text associated with particular types of social occasion (e.g., the news report, the editorial, the cooking recipe)” (Hatim 1997, 217). From a diachronic perspective, genre can be defined as “a type of text whose overall structure and whose grammatical and lexical features have been determined by the contexts in which it is used, and which over time have become institutionalized” (Thornbury 1999, 85). Similarly, for Werlich (1976, 39), “a text type is an idealized norm of distinctive text structuring which serves as a matrix of rules and elements for the encoder when responding linguistically to specific aspects of his experience”.

However, according to Rogers (2000, 11) “genre is a multifaceted concept for which an absolute definition eludes us”. Rogers warns that “the notion of genre and sub-genre, just as the notion of domain and sub-domain, are slippery customers” (2000, 9). Likewise, “the very attempt at establishing a fixed taxonomy of genres may be illusory, given the variety of applications to which genre is relevant and from which its central notion of function is hard to separate” (2000, 11). For Rogers, genre is a system of text classification and the link between terminology and genre is inherent and even historical, as it goes back to documentation and its techniques.

Humbley (1997) and Holzem (1999) explain the obvious relationship between terminology and documentation or information retrieval, as they have “much in common in both the aims of these two activities (producing sets of words enabling a user to access information), their methods (culling texts to extract these words) and their experimentation (with computer-assisted extraction)” (Humbley 1997, 18). Even though Humbley observes as early as 1997 that between the two sub-fields “there has been little cross-fertilization”, with recent projects

such as TermITH¹ (Jacques and Tutin 2015), Terminology and Documentation are fields that seem less separated than in the past. Moreover, the relation between Terminology and Documentation is bridged also by the notion of domain, and hence overlaps with the objectives pursued in LSP studies.

According to Péry-Woodley (2000), the categories of genre correspond to the textual competences of language users. That is, genre constitutes an intuitive category which enables language users to identify, interpret, and produce documents. Moreover, genres are developed by a discourse community in accordance with the main and shared common objectives of the community. For instance, in a scientific setting, the following genres can be distinguished: journal article, conference paper, abstract, book chapter, book, book of abstracts, monograph, PhD or thesis, calls (call for papers, projects, and contributions), experimental protocols, etc. However, the different scientific genres may differ (cf. Bordet 2019) or not all be productive in the same way, when comparing the genres in different disciplines, in Life Science and in History for example. In a university setting, important genres are the following: dissertation, essay, lecture notes, etc.

In a professional setting, genre takes the form of working documents, such as minutes, reports, reference manuals, directories, etc. Each of these genres has specific features, of which some are common to several related genres. For instance, Péry-Woodley (2000, 162) explains that the minutes of a meeting are written in-house documents which belong to the institution and are oriented toward specific target readers, present and absent from the meeting, who share common specialized knowledge with the writer; they are informative and factual and their communicative aim is informative and instructive. Minutes and reports share at least two features: common specialized knowledge and their communicative aim. It should be recalled that communicative purpose is, according to Swales (1990, 10), the prototypical criterion for genre identity.

Moreover, genres are not only domain or setting-dependent, but also culture and language-dependent. A good example is offered by the university setting. In English-speaking countries, one of the central genres is the essay, while in French-speaking countries, this place is occupied by the *dissertation* and the *commentaire*. The French *dissertation* and *commentaire* are both variants of the essay but do not contain the same structural or lexico-grammatical characteristics as the essay. In addition, what is termed *dissertation* in anglophone culture is called a *mémoire* in French culture.

1. TermITH project aims to develop an access to information in various documents from human sciences through indexing based on the terms recognition, disambiguation and analysis (<http://web.atilf.fr/ressources/termith/ressources.php> and <https://anr.fr/Projet-ANR-12-CORD-0029>). Accessed September 2020.

Furthermore, cultural differences in the same genre have been observed by many authors. Bhatia (2002) states that cross-cultural variation in genres is especially manifest in the use of lexico-grammatical and discourse organisational patterns. In a comparative study of scientific articles written by native speakers of English and non-native speakers of English, Fløttum, Dahl and Kinn (2006, 81) show that French researchers use the fewer personal pronouns in comparison with English and Norwegian scholars. For Salager-Meyer, Ariza and Zambrano (2003, 232), who compared articles written in English by Spanish speakers, French speakers and native English speakers in the field of Medicine, French scientists are much more prescriptive, authoritative, and categorical than their English-speaking colleagues. French speakers also use far fewer hedging devices than others.

Another element which contributes to the complexity of genres, as emphasised by Bhatia (2002), is their tendency to evolve and mutually intersect. That is, textual genres are not always pure fixed models but can give rise to mixed models and the established models generally evolve over time. Bhatia (2002) offers a more elaborate view of genre, which in some instances overlaps with the notion of register (see later on in this section). He suggests that genres are intrinsically linked to domains. Bhatia ranks these domain-dependent genre categories as generic systems, which are to be found within professional communities:

System of genres in Law: cases, judgements, ordinances, contracts, agreements etc.
 System of genres in Business: memos, reports, case studies, letters,
 System of genres in public administration: government documents, political communication, news reports, policy statements, international treaties, (...).
 (Bhatia 2002, 10)

He then opposes generic systems of genre with two configurations, which he terms colonies of genres² and mixed and embedded genres:

Although in much of genre analysis, we identify textual artefacts in terms of pure genres, in practice, we often find them in mixed or embedded forms, either because they are designed to achieve a mix of communicative purposes, (...), or to communicate 'private intentions' within the context of 'socially recognised communicative purposes' (...).
 (Bhatia 2002, 10–11)

Ciapuscio (2007) adopts early typologies proposed in 1970's (e.g., Werlich 1975) which allow us to conceptualise genres and text types as models evolving from more abstract ones to the more specific ones, as illustrated in Figure 1.

2. The existence of 'colonies of genre' is prevalent in the theory of genre: Bergmann and Luckmann (1995) term it *family of genres* while Martin and Rose (2008) call it *genre relations*.

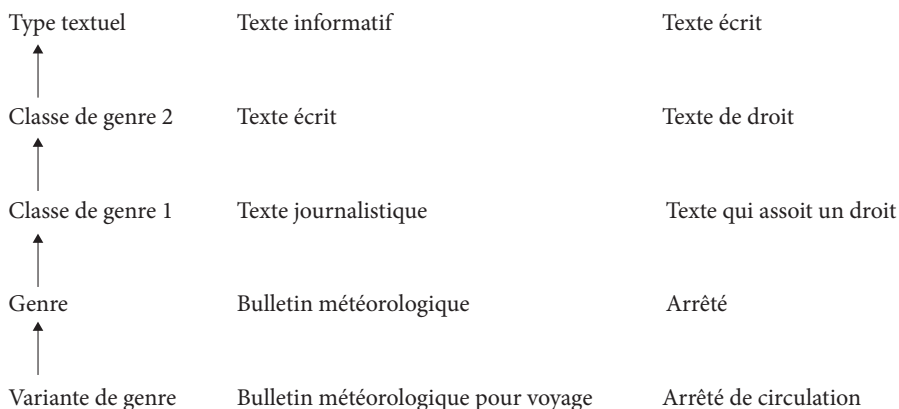


Figure 1. A graded vision of genres, rating from more abstract to more concrete text types (reproduced from Ciapuscio 2007, 408)³

We can thus observe that several factors are responsible for considering genres as fuzzy categories: (i) the profusion of genres; (ii) the possibilities for genres to evolve and multiply by combining features of preexisting genres; (iii) the fact that different discourse communities have specific genres while they can still share some genres with other communities; (iv) the multiple features which define a genre. Moreover, genres are dynamic models which embody the communicative habits of discourse communities and they are culturally bound.

However, there is another important reason that explains the complexity of the category of genre, namely, its relative conceptual synonymy with the notion of register. Bennett (2010) defines register as situations of use and stresses that the terms ‘register’ and ‘genre’ are often interchangeable in corpus studies. Despite their interchangeability, in many instances, these two notions have different, though closely related, meanings. More specifically, genres are specific molds, written or oral, of communication (and hence related to text typology), whereas registers are specific communicative situations in which a set of specific genres are at use.

This distinction is exemplified in Biber’s works (cf. Biber 1995, 1999; Biber and Conrad 2009). According to Biber (1995, 1), registers are “situationally defined varieties” which can be opposed on the basis of text collections. Examples of registers studied by Biber (1999, 8) are conversation, fiction, newspaper language, and academic prose. Moreover, there is generally another meaning attached to the notion of register, in both English and French, which is ‘language level’ and which refers to a

3. Top-down, column 1: Textual type, genre category 2, genre category 1, genre, genre variation; column 2: informative text, written text, journalistic/newspaper text, weather forecast, travel weather forecast; column 3: written text, legal text, text that establishes a right, order, traffic order.

different type of category and set of values, such as literary, formal, neutral, informal, familiar, slang... According to Martin (2002, 57), “register is a pattern of linguistic choices, and genre a pattern of register choice.” Finally, according to Halliday (1985, 29, 38), for whom register is “a variety of language, corresponding to a variety of situation”, the efficient model for genres and register analysis can be achieved “by means of a conceptual framework using the terms *field*, *tenor* and *mode*” (where *field* refers to the subject matter or topic being developed in a particular situation, social context and communicative purposes, *tenor* to roles, cultural values and relationships being enacted in a particular situation, and *mode* to communication context, the type of channel of communication used (e.g., oral, written, visual) and formal communication features). Thus, in Hallidayan vision, a register can be defined as a combination of the field, tenor and mode in a particular situation.

Despite their complexity and the variety of models proposed for their analysis, genres remain essential in all corpus-based and corpus-driven studies. As Rastier (2004, Section I) points out: “[...] l’objet empirique de la linguistique est fait de textes oraux ou écrits, non de mots ou de phrases – qui ne s’observent pas à l’état isolé, et, même quand on les isole, restent toujours relatifs à un genre et un discours.”⁴ Furthermore, Rastier, Cavazza and Abeillé (1994, 194) explain that “le sens est construit en fonction de contraintes globales exercées par le discours [...], le genre du texte et la situation concrète de communication”.⁵ In other words, all empirical studies and observation of a language in use by means of corpora is intrinsically linked to the dimension of genre, represented by the corpus.

More precisely, in terminology research, we cannot overlook four categories or dimensions that correlate whenever we probe the features of terms in context: genre, register, discourse, and domain. According to Holzem (2019) and Rastier (2001), “ainsi le genre assure-il non seulement le lien entre texte et discours mais également entre texte et situation” (Rastier 2001, 229).⁶ Indeed, if we adopt the communicative point of view, genres act as sophisticated models and their role is to facilitate communication by encrypting the purposes and patterns of the communication.

Furthermore, the proposal of Condamines (2003) following Rastier (1998a, 2003) to superpose on textual genres the notion of “interpretative genres” appears

4. [...] the empirical object of Linguistics is made up of oral or written texts, not of words or sentences – which cannot be observed in isolation, and, even when isolated, still relate to a genre and a discourse (our translation).

5. the meaning is constructed on the basis of global constraints imposed by the discourse [...], the genre of the text and the concrete situation of communication (our translation).

6. genres act as the link between the text and discourse and between the text and communicative situation (our translation).

particularly convincing, as it allows the possibility of not only taking into account the conditions of the production of a text but also of its interpretation. The notion of “interpretative genres” highlights the essential role played by a linguist in interpreting language data. According to Condamines (2003), all interpretation of language data is culturally and socially bound by the academic, scientific, and individual position of a linguist. Thus, if we adopt an interpretative linguistic view, genres are sophisticated objects subjected to the interpretation that may differ from one linguist to another. As Holzem and Labiche (2003) remind us, Linguistics belongs not only to human sciences but also to natural sciences – as it looks to understand language use by establishing general rules, models, and the causes behind them – and to interpretative sciences, as it looks to connect the findings with other relevant spheres: social, professional, historical, etc.

4. Terminology work and research through the lens of text genres

This section first discusses typical approaches to text genres in relation to Terminology and then evaluates the place of text genres in the research on terminology.

4.1 Typical approaches to text genres in Terminology

There are two major directions of research in Terminology in respect to textual genres. The first one is the most prominent in Terminology and can be regarded as an indirect approach to genres. Its focus is primarily on terms, which are considered in respect to a specific genre. In the second one, which is centered on genre analysis, terms are taken into account alongside other typical features of the genre with the purpose of characterising the genre as a whole. In our opinion, only the indirect approach, terminology-to-genre oriented studies, falls within the scope of Terminology work and research, whereas the direct, genre-to-terminology studies, is more likely to fall within the scope of genre analysis.

An example of this second direction, which focuses on terminology by first characterizing the typicality of a genre, can be found in Couterut (2017) on procedural texts (e.g., user manuals, recipes). Couterut explains that procedural texts are action-oriented. This has an impact on the way they are written and on the choice of the terms that necessarily characterise this genre.

There is however, a third, direction of research which is concerned with the acquisition of genre competence by learners and language users, who need to communicate in specialised language settings. According to Ciapusio (2007), the comprehension and production of specific genres as skills are generally acquired

through a laborious process which consists in imitating existing models and the productions of specialists. Linguistic approaches and schools, such as the English SFL school (e.g., Martin 2002), the German 'genre linguistics' (*Textsorten*) – one of the most prominent approaches in German Text Linguistics – (cf. Adamzik 2000) and, more generally, Text Linguistics worldwide (e.g., Hyland 2002; Ciapuscio 2007; Benali 2012 among others) have proposed various models to integrate genre into language studies and proficiency.

Our focus here is mainly on the first direction, namely the terminology-to-genre approach, which is more closely related to the various terminological issues. Generally, this type of studies focuses on one specific domain (or a set of related domains) and relies on corpus analysis to explore the discourse in this domain. The selection of sources for corpus design leads to the choice of one or several genres specific of the type of discourse under consideration. An example of this approach is Raus (2013), whose terminological analyses of the circulation of terms used to express the concept of gender equality is carried out from a diachronic perspective in the discourse of various international organizations. Her study not only sheds light on the genres used by various international organizations for communication but reveals the different types of terminological variation encountered: diaphasic, diastratic, diachronic, and diatopic.

Our hypothesis is that in Corpus-based and Textual Terminology, the primary focus is on terms and the secondary focus on texts and corpus, which are regarded as a medium in which terms circulate, occasionally generate variants and synonyms, and sometimes come to life. In the view of Textual Terminology (cf. Bourigault and Slodzian 1999; L'Homme 2004), the term is a construct, which means that it is the result of the analysis made by a terminographer and this analysis takes into account the place occupied by the term in the corpus, the validation by the experts and the objectives pursued by a specific terminographic description. This vision of the relationship between a text and terminology is characteristic of the terminographical perspective. In the terminological perspective, if the term is a construct, it is considered to be the result of the conceptualization process and the subsequent usages made by experts in the texts. In this perspective, its semantic and formal characteristics can be analyzed in texts, in other words, across contexts, with the aim of probing the term behavior, such as its tendency to variation (cf. Pecman 2012b).

If this hypothesis is assumed to be valid (in regard to terms and their relation to texts and corpus), in both corpus-based and textual approach to Terminology with emphases on genres, terms are thus the focus of attention whereas textual genres constitute the dimension within which the terms are identified, observed and analyzed.

4.2 The place of text genres in research on Terminology

Although textual genres are important in research on terminology, there is as yet no study that provides a clear picture of their place in this field. In an effort to resolve this issue, we conducted a survey on research papers in the journal *Terminology*. The results showed that the textual genres used for compiling data in terminological research activities and projects are frequently outlined in research papers reporting on these activities and projects, but in some cases, they can only be implied and, in general, their description is seldom accurate.

In the Appendix, we provide the results of our survey of research articles (21 in total), which were published in *Terminology* from 2017 to 2018, with an emphasis on the place that textual genres occupy in these papers. For each paper, we identified the domain and, when applicable, subdomain, and noted whether the genres used for the study were explicitly identified in the title and/or abstract and/or body of the paper. The information provided in each publication that reflects the credit given to text genres is noted according to its occurrence in the following sections of the paper: title, abstract and body of the paper. Also indicated were the languages under study, year of publication, volume and number of a journal, and the title of the paper.

Figure 2 quantitatively summarizes the data in the Appendix showing that there is a clear tendency to mention the genres in the body of the paper, namely in the

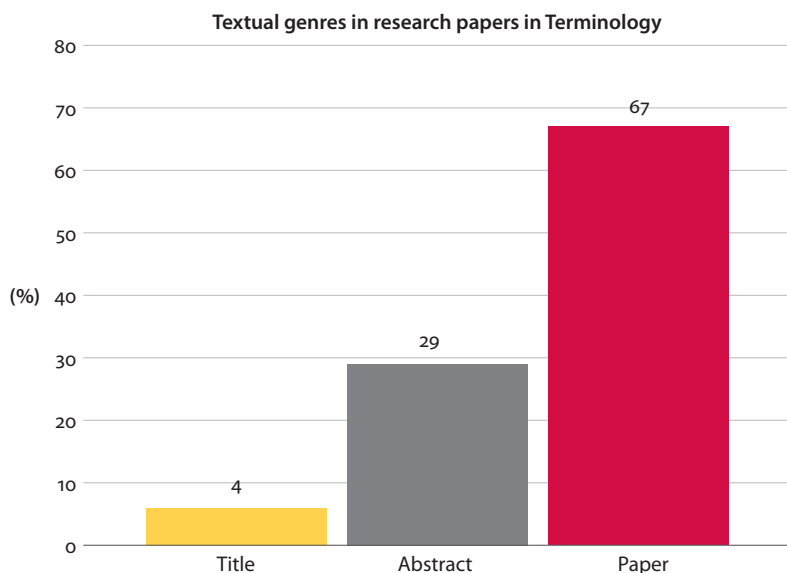


Figure 2. Statistics about the place textual genres occupy in research papers in the journal *Terminology* between 2017 and 2018 in the title, abstract and the body of the paper

parts devoted to methods and corpus design. Genres are sometimes announced in the abstract, but rarely in the title of the paper. In our survey, the sole example is provided by a paper titled “Babel of the senses: on the role of the metaphor and synaesthesia in wine reviews” (see the Appendix, line 19) by Suárez-Toste (2017). A more recent example is López Arroyo and Moreno Pérez’s (2019) study of sales contracts. In rare cases, providing the information about genres is discarded, especially in studies conducted in multidomain perspective (see the Appendix, lines 5 and 6).

The Appendix and Figure 2 show that the greater part of corpus-based research is conducted with an attention to genre types represented by the corpus. However, the results of this survey point to two other constants. First, in the terminology-to-genre direction of research, genres are treated variously. That is, the degree of importance attached to genres is quite variable. Second, genres are rarely described in detail. Instead, they seem to be taken for granted. Furthermore, when a description is provided, if any, it consists in recalling the communicative aims and/or target audience of textual sources used for the study. The focus is thus on sources and corpus size, rather than on the genre and the typicality of the genres. The terminological characteristics studied in a paper are also implicitly presented as one of the typical features of the considered genres.

Figure 3 illustrates a typical credit given to textual genres in terminology research papers. The general tendency consists in briefly presenting the sources used for compiling the corpus, the size of the corpora, the languages and the genre types.

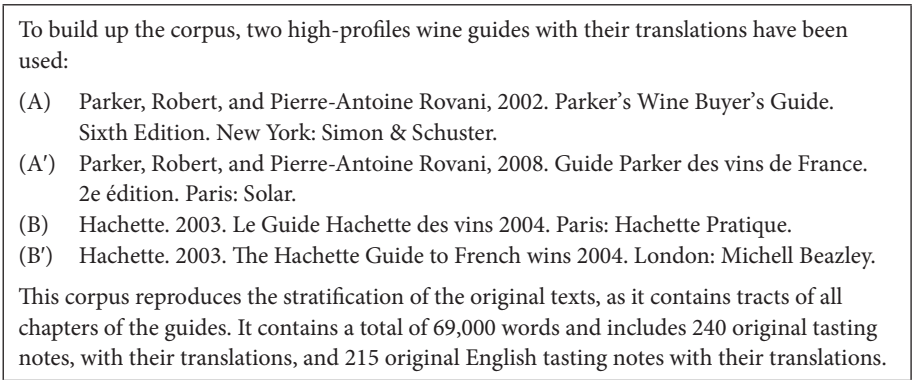


Figure 3. Illustration of a credit given to textual genres in research papers on terminology: Extract from Demaecker (2017, 125, line 20 in the Appendix)

Figure 4 further illustrates the tendency in depicting genre by reference to sources, domain, general context of communication and, most evidently, the corpus design.

BLaRC: A legal corpus of British law reports

BLaRC, the British Law Report Corpus, was designed and compiled with the purpose of identifying and exploring the nature of legal terminology. (...) It is an 8.85 million-word legal corpus of law reports. Law reports could be defined as written collections of judicial decisions made at the UK courts and tribunals.

The reasons to focus on this legal genre to study the linguistic properties of its terminology are varied. To begin with, the UK law system belongs to the realm of common law. Unlike civil or continental law systems, such as those found in most European countries, common law is not codified, that is, it could be stated that, in a way, it is judge-made. Judges interpret the law (found in statutes) and apply the existing case law related to a given case as long as it is similar to that case in its essence. They determine whether a decision made at a higher court should be applicable to a case and how it should be applied. In addition, they must also justify the rationale behind their decision. This is the so-called *ratio decidendi*, which is binding on courts of lower jurisdiction. According to the doctrine of *stare decisis*, which common law revolves around, the judicial decisions (called precedents) made at higher courts such as the Supreme Court or the High Court of England and Wales must be followed by lower courts owing to their binding character, setting a legal precedent within the system, hence their relevance. This is why they are published every year by institutions such as the Incorporated Law Report Council of England and Wales (ILRC) and must be consulted and cited by solicitors, barristers or judges. Therefore, they act as the solid ground which legal practitioners build their arguments on.

Further to their significance within the UK legal system, judicial decisions might be deemed valuable sources of information about the legal lexicon due to the fact that they not only cover all the branches of law, but might also present full embedded sections of other public and private law genres, thus displaying great lexical density and variety.

BLaRC a synchronic, monolingual and specialised collection of 1,228 judicial decisions from the UK courts and tribunals system issued between 2008 and 2010 in raw text format.

Figure 4. Illustration of a credit given to textual genres in research papers on Terminology: Extract from Marín and Rea (2014, 3)

Moreover, we can notice that in the titles, the authors generally announce the domain coverage, suggesting a dominance of domain over genre. For instance, Miyata and Kageura (2018) conduct an analysis with a view to providing information on the terminology of local government, as announced in the title of their paper (line 1 in the Appendix). They identify the domain coverage and the text genres targeted by their analysis: “municipal information such as residential procedures, tax payment and child care” (2018, 156), but the specificities of the genre at scope remain implicit. However, the authors name three websites that they selected for the purpose of collecting textual data.

Finally, research in Terminology is not limited to terms but may include studies on phraseology and various lexical structures (e.g., lexical chunks, complex

noun phrases...) as supplementary features of specialized texts (e.g., Kübler, Mestivier-Volanschi and Pecman 2022; Pecman 2021). These studies can be regarded as common to the sphere of research on terminology, phraseology and LSP. Actually, research in Terminology is adjacent to a series of other fields of study, which are generally corpus-based – phraseology, genre studies, discourse analysis, LSP studies, specialized Translation Studies, etc. – and often a work in Terminology pertains also to one or more of these fields. All these fields of studies are bound by the dimension of genre.

5. Basic principles and methodologies

The hypotheses of studies on textual genres and terminology generally consist in considering that the choice of terms, especially in the case of variants, is related to certain genre types, and consequently conducting research on terms by targeting specific genres may prove a fruitful method for comprehending term usage, term variation and circulation.

Although terms cannot be captured independently of domains, discourse types, registers and genre, there are very few papers theorizing the approaches to Terminology with special emphases on the central role played by the text genres therein. The rare exception is offered by Rogers (2000) in Trosborg (2000). However, there are some theoretical elements that apply to genre in Bourigault and Slodzian's (1999) on the principles of Textual Terminology.

5.1 The obvious link between genre and domain

The first principle of Textual Terminology applied to genre analysis is that the choice of a genre or set of genres is related to the choice of a domain. In other words, there is a relationship between genres and domains, which is due to the fact that each knowledge domain has at its disposal a set of predefined and accepted genres for ensuring communication. The choice of one over the other is thus correlated. Péry-Woodley (2000, 134) refers to the relationship between genre and domain as “genre-domain configuration”. She explains that it involves the intersection of one domain with a certain genre or document type (e.g., reference manual and Computer Science).

The choice of a genre seldom comes after the choice of a domain. In this case, the choice of a genre will open for the researcher a set of possibilities for selecting domains that use the genre for communication. A good example of this is Rogers (1997) on the compounding possibilities of terms (from the perspective

of synonymy and equivalence) in scientific articles (as a genre) and on Genetic Engineering (as a domain or field). This is also the case in Pecman (2012b, 2014) on terminological variation in scientific articles, who focuses on an array of disciplines, namely, Climatology, Geodynamics, Glaciology, Seismology and Volcanology. In order to ensure the data coherence pertaining to a certain genre, the selected domains/disciplines are often closely connected. In the case of Pecman (2012b, 2014), the five disciplines belong all to a single wider domain of Earth and Planetary Sciences (EPS). The choices of genre and domain(s) are essential.

Conversely, in the majority of studies, the choice of a domain governs the selection of genres. In other words, choosing a domain is usually prior to choosing a genre or a set of genres for conducting terminological research. Domain choice thus appears to be the first and foremost dimension in terminology studies. Indeed, there is no research in Terminology outside a specific domain. In domain-to-genre research, the choice of genre comes after the choice of domain. This configuration is explored by Jacques (2000) in her study on the reduction of complex terminological units. She describes her corpus according to the domain and genres, which are defined as corresponding to specific communicative and discursive targets of the expert community:

Ces textes sont donc situés par rapport : 1. à un domaine de connaissances : la gestion des déplacements dans une agglomération ; 2. à certaines visées discursives qui donnent forme aux documents de travail : cahiers des charges, règlement d'appel d'offres, glossaire et manuel d'utilisation d'un système informatique, notes de service, circulaires et projet de partenariat [...]. (Jacques 2000, 100)⁷

The majority of studies in Terminology focus on one domain or specialised discourse (and less frequently on two or several by adopting a contrastive perspective). The choice of a genre or set of genres is always relevant, as discussed in the previous section. However, as genre selection is dependent on domain selection, the consideration of genres in some studies is implicit. The results of such studies remain dependant on the characteristics of the genres. This is especially true when terminology research is based on a corpus, which has been carefully compiled to include the relevant genre types.

Rogers (2000) explains that there are different methods that can be used to compile a corpus in respect to genre-domain configuration. Figures 5 and 6 illustrate two of these methods.

7. The texts are situated in relation to: 1. a field of knowledge: ex. travel management in built up area around a large city; 2. discursive aims which define the working documents: specifications, call for tender regulations, glossary and user manual for a computer system, memoranda, prospectus and partnership project [...] (our translation).

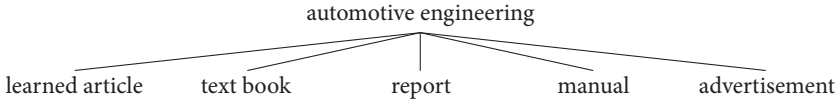


Figure 5. A simple LSP corpus design with two attributes: Domain and genre (reproduced from Rogers 2000, 9)

It goes without saying that the number of domains has experienced exponential growth with 300 domains estimated in the 1980s and 30,000 in the 1990s (see Rogers 2000, 8). Even though new genres occasionally appear, it is more frequent for existing genres to evolve internally. On the other hand, as underlined by Kuteeva and Negretti (2016), the boundaries between disciplines have become increasingly blurred. This situation may create an obstacle for studying the terminology of a single discipline.

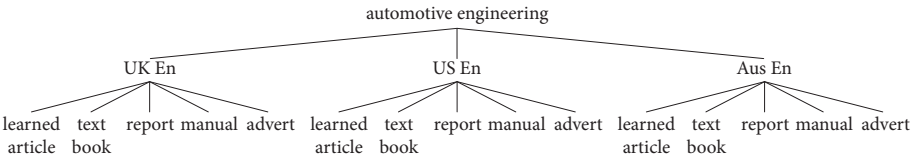


Figure 6. Corpus organized hierarchically: Domain, language variety, and genre (reproduced from Rogers 2000, 10)

5.2 The link between genre and text

The second principle underlying terminology studies applied to genre analysis is that the link between terminology and genre is provided by the text, as Rogers (2000, 3) observes that the notion of text links terminology and genre. It should be recalled that in the view of Textual Terminology, the text is the point of entry for the description of terms.

This second principle is responsible for the integration of corpus-based methods for the analysis of genres and terminology. There are two sub-methods for corpus analysis: linear and diagonal. The linear method entails going back and forth from the corpus to the text in order to take into account the linear nature of the text and its unicity when analysing the terminology. The diagonal method involves browsing the contexts and concordances in the corpora in order to bypass the linear nature of the texts. This second method is more likely to be related to corpus-based approaches to Terminology than to Textual Terminology. In the textual approach, the focus is on the text and the terms are never analysed isolated outside of their entire textual context (in contrast to a Corpus Linguistics approach which consists of the tendency to take into account the immediate contexts).

Condamines (2018b) gives a moderately different perspective of the methods used in Textual Terminology. Her view is text-centered rather than genre-centered, regardless of how contiguous the two dimensions may be. According to Condamines (2018b, 1), “terms and the relationships (spotted initially in order to build terminological knowledge bases) become a way to enter the texts. Terms act as pivots, and the interpretation of their contexts is crucial to the aim of the study.” She explains that using texts and corpora is essential for building terminological knowledge bases, by means of term networks analysis (Condamines 2002, 2018b). The method adopted for constructing terminological networks greatly depends on the domain but also on the target applications of the study. Generally, the methods adopted consist in using linguistic markers and distributional analysis to define terms and their conceptual relations.

5.3 The link between genre and corpus

Genres are intrinsically linked to domains, texts, and the corpus design. Work on text genres in Terminology is mediated by the corpus, and relies on the general methodology for corpus design, exploration, and exploitation. Working with corpora in Terminology is discussed in Bourigault and Slodzian (1999). The research group, Terminologie et Intelligence Artificielle (TIA), created by Didier Bourigault and Anne Condamines, established the methodology for working with corpora from a terminographic perspective (see Figure 7) (cf. Campo 2012, 173).

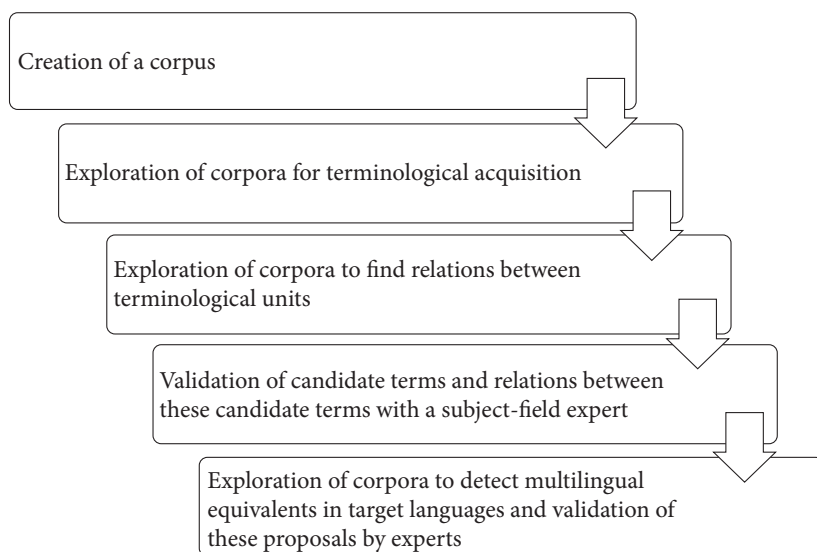


Figure 7. Stages of a corpus-based textual approach to Terminology conducted in a terminographic perspective (after Bourigault and Slodzian 1999)

This methodology designed by Bourigault and Slodzian (1999) can be extrapolated to the following framework for integrating genre in research conducted in a terminological perspective. The dimension of genre remains implicit in Bourigault and Slodzian’s scheme. Figure 8 proposes the characteristic steps in working with corpora in a terminological perspective, by making explicit the role of text genres.

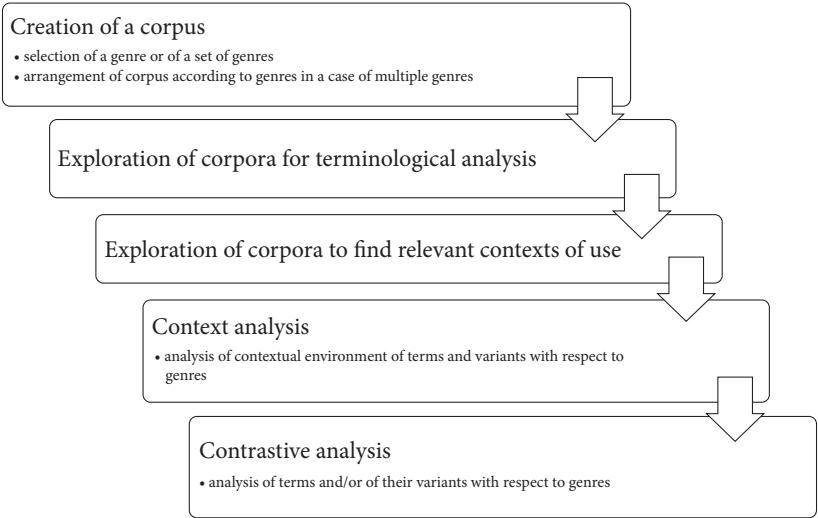


Figure 8. Stages in corpus-based and textual approach to Terminology with emphasis on text genres

The two methods are comparable in many respects but they differ in their purpose and aims. In the terminographic perspective, the aim is to establish equivalences between languages and create reusable multilingual terminological resources. Thus, text genres are implicitly considered. Focusing on text genres and adopting a terminological perspective (as illustrated in Figure 8) foregrounds the understanding of the way terms behave and their dynamics in respect to genres.

6. Impact of textual genres on Theoretical and Applied Terminology

From a linguistic point of view, one might expect that the major outcome of studies on terminology in relation to textual genres would be to shed light on the terminological characteristics of one genre in comparison to another within the same domain. However, such contrastive analyses are actually rare. Most frequently, the outcome of terminological studies on textual genres consists in showing the following:

1. The important role of genre in interpreting terminological data;
2. The important role of terminology in characterising and interpreting the characteristics of a certain genre;
3. The extent to which terminology appears as a classifier of a genre;
4. The implications of terminological variation within a genre or across genres;
5. New perspectives resulting from computer-based technologies, most notably for the computerised management of textual corpora and genres.

The last two outcomes are not specific to studies on terminology in relation to textual genres but also to studies on corpus-based approaches to Terminology and to studies on terminological variation (cf. Picton and Dury 2017).

From an epistemological perspective, Corpus-based and Textual Terminology are closely linked to the majority of theoretical approaches as the text has become the most important medium for collecting terms. Thus knowledge-based terminology, studies on terminological variation and cognitive studies are all to some extent corpus and text-based. In fact, the corpus-based approach confirms the importance of working with authentic data, whereas the textual approach confirms the importance of a text as a communicative unit. Textual Terminology reminds us that, alongside the corpus as a collection of texts, it is important to bear in mind that the text is a unit of meaning, and that terms need to be considered also within this specific unit.

Moreover, the category of genre has played an important role in the development of Terminology into a discipline. Terminology long seemed to be a means rather than a discipline (cf. Sager 1990; Humbley 2017). It now occupies a well-defined place in institutions, universities, to some extent in private companies and in society. It is generally integrated into curricula on Translation Studies, where it is taught by specialists in Terminology. Terminology is also relevant in all disciplines since it is essential for experts to reflect on the terms used.

The purpose of this chapter was to bring to light some elements pertaining to the role of genres in Terminology. The issues raised by genres, texts and corpus have not only improved our methods for analysing and working with terminology but have also enhanced our knowledge of terminology, and by the same token of the way that this knowledge is conceptualised and designated by terms. The corpus-based and textual approaches offer methods to observe the behaviour of terms in texts and thus better understand the processes of their formation, how these processes can influence knowledge construction, hamper or facilitate comprehension and translation, and ultimately shed light on the terminological characteristics of specialised textual genres. As terms are multidimensional units, being at once units of knowledge, units of language, and units of communication, their analysis in relation to specific genres steadily brings forth the cognitive, linguistic

and communicative characteristics of the genres. As genres are complex constructs, we can argue that the terminological approach helps understanding some of the subtle features of genres, and their implications for communication.

7. Conclusion

This chapter has shown that in both research and practice in Terminology, the category of genre is intrinsically linked to the categories of domain, and subsequently to text and corpus. Text genres are coded models according to which language users identify, interpret, and produce documents. However, the notion of genre is a fuzzy category because of its complexity, varieties and overlapping with the notion of register. In Terminology, genres generally refer to text types and are closely linked to textual sources used for compiling corpora. Although the importance of text genres in terminology practice and research is widely acknowledged, in some instance genres are implied rather than characterized because they are considered to be universally known.

We also showed that Corpus-based Terminology as well as Textual Terminology offer new methods of analysing terms by allowing terminologists and linguists to explore them in relation to textual genesis and specific features of textual genres. In other words, corpus-based and textual approaches to Terminology offer a framework of specific tools and methods for prospecting terminology from any of the theoretical standpoints that characterize Terminology as a discipline. Moreover, given the complexity of genre, terminological analysis applied to genre made it possible to isolate and comprehend crucial elements about the latter. However, further studies would be necessary in order to devise models for systematically taking into account genre internal and external features in studies on terminology from the perspective of corpus design and data interpretation.

From this overview of the role played by text genres in terminology research and practice, the conclusion is that our knowledge and vision of terms greatly depend on the way that they are studied, on the perspectives adopted, on the methods and tools used and on what is being measured and/or proven. In the light of the evolution of Terminology as a discipline, capable of developing its own methods for conducting research and managing terms, it can be predicted that approaches to Terminology will continue to diversify in order to respond to new challenges and needs. If terms are the embodiment of concepts that crystallise our expert knowledge about various subject fields, it is highly probable that analysing terms through the lens of corpus, text, and genre analysis will remain a valuable tool for investigating lexical, semantic, pragmatic, and cognitive issues pertaining to the use of terms in text types and genres.

Appendix

Domain	Subdomain	Information on genre in the TITLE	Information on genre in the ABSTRACT	Information on genre in the BODY OF THE PAPER	Languages	Year	Vol(n)	Paper title	
1	administration	local government	Ø	Ø	residential procedures, tax payment, child care	Japanese, English	2018	24(2)	<i>Building controlled bilingual terminologies for the municipal domain and evaluating them using a coverage estimation approach</i>
2	medicine	traditional Chinese medicine	Ø	dictionaries, journals and official websites	“Xie’s series of articles (1997–2000), Li’s dictionary (2002) and the official website of the Committee for terms in TCM (CTTCM)”	English, Chinese	2018	24(2)	<i>English translation of long Traditional Chinese Medicine terms</i>
3	selling	selling cheese	Ø	Ø	promotional texts selling cheese from promotional websites of cheese manufacturers and retailers, and from websites promoting cheese in general	English	2018	24(2)	<i>Selling cheese online</i>

Domain	Subdomain	Information on genre in the TITLE	Information on genre in the ABSTRACT	Information on genre in the BODY OF THE PAPER	Languages	Year	Vol(n)	Paper title
4	linguistics	Ø	14 EFL/ESL pedagogic grammar textbooks	“three major books (Berry 2010; Leech 2006; Spears 1991) along with as many as ten other sources – such as Active Grammar series (Davis and Rimmer 2011) and An Introduction to English Grammar (Greenbaum and Nelson 2002)”	English	2018	24(2)	<i>Towards finding a difficulty index for English grammatical terminology</i>
5	humanities and social sciences	Ø	Ø	Ø		2018	24(2)	<i>Conceptualization and theorization of terminology translation in humanities and social sciences</i>
6	multidomain	Ø	Ø	Ø	Japanese	2018	24(1)	<i>NWJC2Vec. Word embedding dataset from ‘NINJAL Web Japanese Corpus’.</i>
7	mathematics	Ø	Ø	5 English high school mathematics textbooks used in Ontario, Canada covering grades 9 to 12 and including topics on functions, calculus and vectors	English	2018	24(1)	<i>Distributed specificity for automatic terminology extraction</i>

Domain	Subdomain	Information on genre in the TITLE	Information on genre in the ABSTRACT	Information on genre in the BODY OF THE PAPER	Languages	Year	Vol(n)	Paper title	
8	medicine	clinical domain	Ø	clinical records	Electronic Health Records (EHR) from the endocrinological ward of Belgian hospital, containing medical history of patients	Flemish Dutch	2018	24(1)	Clinical sublanguages. Vocabulary structure and its impact on term weighting.
9	multidomain	medicine, art, literature, economy, etc.	Ø	narrative clinical records and biomedical articles	patient records, book chapters and journal papers, journalistic books, etc.	Polish	2018	24(1)	Recognition of irrelevant phrases in automatically extracted lists of domain terms
10	medicine	clinical domain, phenotypes	Ø	narrative clinical records and biomedical articles	Ø	English	2018	24(1)	HYPHEN. A flexible, hybrid method to map phenotype concept mentions to terminological resources.
11	multidomain	economics, health and social service	Ø	Ø	IATE's economic and health glossaries and various corpora	English, Spanish, French, Catalan	2018	24(1)	Improving term candidates selection using terminological tokens
12	medicine	AIDS, HIV	Ø	Ø	“data for this study were collected through evaluative survey and language-specific workshops”	Hausa, Igbo, Yoruba	2017	23(2)	The modernisation of HIV and AIDS' nomenclatures in Nigeria's major languages

Domain	Subdomain	Information on genre in the TITLE	Information on genre in the ABSTRACT	Information on genre in the BODY OF THE PAPER	Languages	Year	Vol(n)	Paper title
13	medicine	Ø	Ø	survey among students, dental literature	Croatian	2017	23(2)	<i>Assessing the influence of the English language on the professional vocabulary of Croatian dental students by analysing their word choice for the translation of medical/dental terms</i>
14	informatics	Ø	Ø	"For each language, we chose two programming manuals, one on Linux and the other on Phyton, and one PC maintenance manual."	Arabic, English, French	2017	23(2)	<i>Automatic extraction of specialized verbal units</i>
15	food	Ø	"general language corpus", "literary works of authors such as Sandra Cisneros, Najat El Hachmi, Chimamanda Adichie, and others"	all the novels used for the study are presented in detail	English	2017	23(1)	<i>Food terminology as a system of cultural communication</i>

Domain	Subdomain	Information on genre in the TITLE	Information on genre in the ABSTRACT	Information on genre in the BODY OF THE PAPER	Languages	Year	Vol(n)	Paper title
16 sensory experiences	sight, smell, taste	Ø	tasting notes	“tasting notes (...) written by the most authoritative critics in the field (...)”; the publications are named	English, Spanish	2017	23(1)	<i>From the glass through the nose and the mouth</i>
17 sensory experiences	taste	Ø	lexicons, dictionaries, web-posting, and elsewhere	“we collected data from the lexicons, dictionaries, web-posting”; the sources are named.	Korean	2017	23(1)	<i>Multifaceted gustation</i>
18 sensory experiences	wine	Ø	Ø	enquiry among various “participants with different types of expertise in wine”		2017	23(1)	<i>How words for sensory experiences become terms</i>
19 sensory experiences	wine	wine reviews	“English and Spanish wine reviews (...) that best exemplify the many of the genre’s shortcomings.”	“a text corpus (...) drawn from some of the most authoritative wine publications (...)”; the publications are named	English, Spanish	2017	23(1)	<i>Babel of the senses: on the role of the metaphor and synesthesia in wine reviews</i>
20 sensory experiences	taste, drinks, wine	Ø	Ø	“two high-profile wine guides and their translations”; the guides are explicitly named and referenced	English, French	2017	23(1)	<i>Wine-tasting metaphors and their translation</i>
21 sensory experiences	marketing, smoothies	Ø	Ø	various textual sources explicitly named and referenced	English	2017	23(1)	<i>Verbalizing sensory experience for marketing success</i>

Knowledge patterns in corpora

Elizabeth Marshman

University of Ottawa

Knowledge patterns (KPs) have evolved as tools for detecting, extracting, and classifying occurrences of conceptual and semantic relations expressed in texts. KPs offer many strengths, including clear identification of relations present. The extraction of relation-containing contexts, often referred to as *knowledge-rich*, allows users to observe the items participating in the relation and nuances conveyed by other textual elements. Nevertheless, designing and implementing pattern-based approaches is challenging, in part because of patterns' linguistic (i.e., lexical, grammatical) or paralinguistic nature. KPs must be selected and represented with care for a given use, and their performance may vary. In this chapter, we will explore some benefits and challenges of pattern-based approaches to corpus analysis, and some prospects for future applications.

Keywords: knowledge patterns, conceptual relations, semantic relations, corpora, knowledge-rich contexts, semi-automatic knowledge extraction, terminology, ontology

1. Key concepts and definitions

Interest in knowledge patterns in corpora evolved from growth in the use of corpora for terminology research (cf. Meyer and Mackintosh 1996), which in turn facilitated the development of a number of theoretical approaches in Terminology that highlight conceptual and/or semantic relations in the acquisition, description, and management of concepts and terms (cf. Nuopponen, this volume). These approaches include lexico-semantic terminology (cf. L'Homme in this volume) and frame-based terminology (cf. Faber, this volume). Relations are valued as central elements in users' acquisition and understanding of new concepts and domains (e.g., Faber, León-Araúz and Reimerink 2014; San Martín, Trekker and León-Araúz 2020). This makes them critical for both creators and users of terminological resources, and one of the underpinnings of terminological description in a wide variety of resources (whether or not they are explicitly highlighted).

Pattern-based approaches to corpus analysis are based on the assumption that there are recurrent and predictable linguistic cues that indicate specific types of information in texts, and that may serve to locate and extract excerpts of these texts – and thus the knowledge they convey – for use in applications such as terminography (Meyer et al. 1999, 257).

These information-containing excerpts are referred to as *knowledge-rich contexts* (KRCs). Meyer (2001, 281) defines a KRC as “indicating at least one item of domain knowledge that could be useful for conceptual analysis,” i.e., at least one conceptual attribute or relation. The linguistic cues, or *knowledge patterns*, were known under a variety of names in their early incarnations, for example, Ahmad and Fulford’s (1992) *knowledge probes*, Bowden, Halstead and Rose’s (1996) *relation markers*, and Pearson’s (1998) *defining expositives*. They drew early interest as tools to facilitate and accelerate the extraction of KRCs from text corpora because as Meyer (2001, 281) observes, “a corpus is only as useful as the tools available for extracting... data.”

KPs are most often used to identify KRCs expressing relations, which may be viewed through a conceptual lens (e.g., most commonly generic-specific and part-whole, along with cause-effect and object-function), or a semantic one (e.g., synonymy, equivalence). They not only signal that a relation is present, but also generally identify the relation type, and point to the items (usually terms and by extension concepts) linked by the relation.

Meyer (2001, 290–291) describes three main KP types:

- Paralinguistic, including punctuation and typography, such as in *knowledge-rich contexts (KRCs) can be spotted in corpora*, where parentheses indicate synonymy between the term and initialism;
- Grammatical, involving sequences of parts of speech, such as in *the software searches the corpus*, where the sequence of noun (*software*) and verb (*to search*) indicates an object-function relation;
- Lexical, typically in the form of “triple” structures consisting of a lexical relation marker (i.e., a word or phrase) linking two items and identifying the relation between them, such as in *a concordancer is a type of software*, where the lexical marker *is a type of* links the specific term *concordancer* and its generic term, *software*.¹

1. Reimerink, León-Araúz and Faber (2016), in the EcoLexicon project (cf. Faber, this volume), identify an additional type of knowledge pattern, the *visual knowledge pattern* (VPK), corresponding to elements such as labels or other textual elements, colours and arrows in images such as diagrams and flow charts. VPKs may indicate, for example, the parts of a process or the functions of an object. They will not be explored in depth in this chapter, although their potential in resources with a strong visual/image-based component is an interesting area for further investigation.

This chapter focuses mainly on lexical knowledge patterns.

Table 1 shows a sample of English markers that indicate commonly investigated relations.

Table 1. Examples of lexical relation markers²

Relation	Lexical marker ² (typically occurring in pattern form Term + Marker + Term)
Generic-specific	<i>is a, is a type of, is a kind of, includes, encompasses, such as, and other</i>
Part-whole	<i>is a part of, is composed of, comprises</i>
Cause-effect	<i>causes, caused by, leading to, is a result of, triggers, is responsible for, affects</i>
Object-function	<i>used to, needed for, designed to, provide</i>
Synonymy	<i>also called, also known as</i>

In this chapter, Section 2 will discuss KP applications and their impacts on terminology work, as well as some challenges of pattern-based approaches. Section 3 will describe some preparatory tasks required for these applications and their associated challenges, and Section 4 will provide an overview of some evaluation strategies for such applications. Finally, some ideas for future developments are presented in Section 5.

2. Exploration and application of knowledge patterns

This section offers a portrait of the range of projects, languages, subject fields, and relations in which KPs have been investigated. We also explore some of the ways in which KPs may be used, including for corpus analysis, corpus-building, and the enrichment of terminology resources.

2.1 Exploration of knowledge patterns

KPs have been researched in a variety of languages, with English still the most common, but a growing representation of other languages, as shown in Table 2.

Similarly, scholars have studied KPs in a range of specialized subject fields, as shown in Table 3. Most projects have tended to concentrate on a single field, although some (e.g., Jacques and Aussenac-Gilles 2006; Marshman, L'Homme and Surtees 2008a; San Martín, Trekker and León-Araúz 2020) have included two

2. These markers have been collected from Meyer et al. (1999); Meyer (2001); Condamines and Rebeyrolle (2001); Marshman (2007a), and Faber, León-Araúz and Reimerink (2014).

or more (often with a comparative orientation and/or the aim of identifying patterns likely to be domain-independent, that is, relevant to a variety of fields, cf. Section 3.1.1 below).

Table 2. Languages of KPs investigated

Language	Examples of studies
English	Hearst 1992; Ahmad and Fulford 1992; Meyer et al. 1999; Gillam, Tariq and Ahmad 2005; León-Araúz, San Martín and Faber 2016
French	Garcia 1997; Rebeyrolle 2000; Séguéla 2001; Marshman, Morgan and Meyer 2002; Malaisé, Zweigenbaum and Bachimont 2005; San Martín, Trekker and León-Araúz 2020
Spanish	Alarcón, Bach and Sierra 2007; Sierra et al. 2008; Soler and Alcina 2008; Aguado-de-Cea, Álvarez de Mon and Montiel-Ponsoda 2009
Catalan	Feliu 2004
Danish	Weilgaard Christensen 2019
Dutch	Lefever, Van de Kauter and Hoste 2014
Russian	Schumann 2012

Table 3. Subject fields investigated

Subject field	Examples of studies
Life Sciences	Meyer et al. 1999; Condamines 2002; Marshman 2008; Aguado-de-Cea, Álvarez de Mon and Montiel-Ponsoda 2009
Psychology	Malaisé, Zweigenbaum and Bachimont 2005; Marshman, L’Homme and Surtees 2008a
Computing	Condamines and Rebeyrolle 2001; Marshman, Morgan and Meyer 2002
Environment and Earth Sciences	Meyer 2001; Barrière 2001; Faber, León-Araúz and Reimerink 2014; Josselin-Leray et al. 2014
Engineering and Technology	Ahmad and Fulford 1992; Garcia 1997; Condamines and Rebeyrolle 2001; Gillam, Tariq and Ahmad 2005; Weilgaard Christensen 2019

As noted previously, KPs for a number of relations have been identified. Among the most commonly studied are the hierarchical relations generic-specific (or hyperonymy) and part-whole (meronymy), as these are most often used in the creation of concept systems in Terminology (cf. Nuopponen, this volume). However, a growing range of relatively domain-independent non-hierarchical relations have also been considered, including cause-effect, object-function or purpose, and association. From a semantic perspective, relations such as synonymy have also drawn attention.

Table 4 gives some examples. In addition, a number of domain-linked relations (e.g., symptom in the field of Medicine) (e.g., Meyer et al. 1999) may be identified and studied for specific purposes.

Moreover, many relations may be internally classified to highlight different sub-types (e.g., Bowden, Halstead and Rose 1996 for part-whole; Barrière 2001 for cause-effect), and some classifications can achieve a very fine level of granularity, depending on the specific needs of the application. Faber, León-Araúz and Reimerink (2014), for example, devised a set of conceptual relations to meet the specific needs of the domain (Environment) and concept classes they connect, as well as the specific properties (e.g., transitivity) required for optimal management and representation of relation data in the EcoLexicon.

Table 4. Relation types studied

Relation type	Examples of studies
Generic-specific	Hearst 1992; Ahmad and Fulford 1992; Meyer 2001; Marshman, Gariépy and Harms 2012; Gillam, Tariq and Ahmad 2005; Aguado-de-Cea, Álvarez de Mon and Montiel-Ponsoda 2009; León-Araúz, San Martín and Faber 2016
Part-whole	Ahmad and Fulford 1992; Bowden, Halstead and Rose 1996; Meyer 2001; Marshman, Morgan and Meyer 2002; León-Araúz, San Martín and Faber 2016
Cause-effect	Ahmad and Fulford 1992; Garcia 1997; Meyer et al. 1999; Barrière 2001; Marshman 2007a; León-Araúz, San Martín and Faber 2016
Object-function/Purpose	Meyer 2001; Marshman, Morgan and Meyer 2002; León-Araúz, San Martín and Faber 2016
Association	Feliu 2004; Marshman 2007a
Synonymy	Ahmad and Fulford 1992; Meyer et al. 1999

2.2 Applications of knowledge patterns

One of the first intended applications of KPs to extract KRCs from corpora was the creation of terminology resources and ontologies. In a description of traditional term-bank centred work, Meyer (2001, 282) highlights three possible uses of KRCs by terminographers: (i) as definitions in term records; (ii) as starting points for formulating such definition; and (iii) as a means of acquiring domain knowledge. In such traditional models of terminology resources, relations were very often used in research stages and present implicitly in textual supports, but were rarely specifically listed in entries. However, in recent decades, increasing attention has been paid to making relations explicit, and such information is now highlighted

in resources such as terminological knowledge bases (TKBs) (e.g., Meyer 2001; Condamines and Rebeyrolle 2001; Faber, León- Araúz and Reimerink 2014) and ontologies (e.g., Gillam, Tariq and Ahmad 2005; Aussenac-Gilles, Després and Szulman 2008; Aguado-de-Cea, Álvarez de Mon and Montiel-Ponsoda 2009).

In recent years, the automated construction and enrichment of semantic resources and ontologies has drawn increasing attention. Given the variety of applications and the challenges of using KPs to extract KRCs (cf. Section 2.2.2.1 below), full automation is nevertheless still seen as the exception rather than the rule (Aussenac-Gilles and Condamines 2012). Condamines and Rebeyrolle (2001, 127–128), for example, note that a TKB (in their case, specifically a corpus-based terminological knowledge base, or CTKB) may in fact require a much less formal approach than one intended for the development of a formal ontology or similarly application-driven model, as it will need to reflect the goals motivating its construction in its form and content.

2.2.1 *Alternatives and advantages*

For the extraction of relations from corpora, KP-based approaches may be considered either as alternatives to formal or statistical techniques, or as complements to them in hybrid approaches (e.g., Gillam, Tariq and Ahmad 2005; Aguado-de-Cea, Álvarez de Mon and Montiel-Ponsoda 2009). Some relations may be identified based on the analysis of term structures (e.g., shared morphemes or term heads), while others may be revealed by distributional analysis (e.g., Auger and Barrière 2008; Aussenac-Gilles and Jacques 2008). Formal and statistical techniques have a number of advantages, in that they are relatively language-independent, transposable with minimal time and effort from one language to another (and one corpus to another). As discussed in Section 3.1.1, this is not the case for pattern-based approaches. On the other hand, particularly distributional approaches (cf. Bertels, this volume) may provide very little indication of the specific nature of the relationship in a given context, and require human interpretation before the results can be used for many applications. They are also dependent on the presence of multiple occurrences of relations in corpora.

Another approach, in documents with standard components such as Wikipedia pages, uses the structure of the documents (e.g., Kamel et al. 2017). Of course, this technique relies not only on the availability of highly structured documents, but also the uniformity of the structure, two limiting factors for most text types of interest for corpus-building. The rise of machine-learning techniques is leading to the development of new approaches, although these still often require extensive annotation of training corpora as a starting point, an investment that may be a challenge for developing new applications (Kamel et al. 2017).

Because of their advantages, KPs have continued to be used extensively for corpus analysis, but also for their potential for building useful corpora and for identifying useful contexts for consultation by language professionals such as translators.

2.2.2 *Use for corpus analysis*

Meyer (2001, 288) describes what she refers to as a *term-oriented* approach to using knowledge-extraction tools, which involves the preliminary identification of a set of terms or candidate terms of interest (e.g., using a term extractor or through manual scanning) and then using a knowledge extraction tool that searches for a term of interest in proximity to one or more relation markers. She also posits (299) the potential for a *relation-oriented* approach that could extract term pairs linked by a chosen relationship. This approach has been used for example by Malaisé, Zweigenbaum and Bachimont (2005) to detect definitions in texts.

In either case, searching corpora for occurrences of KPs requires some kind of corpus-analysis tool and a formal representation of KPs that can be used in search queries (cf. Section 3.2). Researchers have used both “off-the-shelf” and custom-built tools for this purpose.

KPs represented using simple character strings or regular expressions may be fairly easily implemented in simple key-word-in-context (KWIC) concordancers (e.g., WordSmith Tools,³ AntConc⁴). These tools allow users to locate occurrences of KPs, often in combination with a term of interest, to count frequencies, to view the occurrences in a short or extended context, and often to sort the occurrences to help to bring out regularities (e.g., the other term or terms to which the search term is linked by the lexical marker). (Of course, as a variety of patterns may indicate a given relation, it is necessary to generate queries for a list of KPs, sometimes referred to as a *pattern set*.) Users can then evaluate the results, choose what information to retain for their purposes, and convert it to the required format for use (e.g., copy a context or definition to add as a textual support on a term record, enter it into a database of relations that may be visualized within a TKB interface, or convert it to a formal representation used to build an ontology).

To facilitate the application of patterns to extract relation occurrences and the subsequent use of the relations discovered, a number of researchers have developed their own customized tools. For reasons of space, unfortunately only a few

3. WordSmith Tools. Lexical Analysis Software/Oxford University Press (<https://www.lexically.net/wordsmith/>). Accessed 23 December 2020.

4. Anthony, Laurent. AntConc (<https://www.laurenceanthony.net/software/antconc/>). Accessed 23 December 2020.

can be named here. In the COGNITERM project (Bowker 1993; Meyer 2001), the DockMan toolset was used to manage and analyze a corpus to feed the representation managed in the Conceptually Oriented Description Environment (CODE). Garcia (1997) developed the COATIS system for the identification and extraction of cause-effect relations. Séguéla (2001) developed the CAMÉLÉON system, later used by Jacques and Aussenac-Gilles (2006) and integrated into the approach used in the TERMINAE platform (Aussenac-Gilles, Després and Szulman 2008). Barrière (2004) produced a system called SeRT (Semantic Relations in Text) that combined term extraction, KP-based semantic relation searching, and relation visualization functions for a variety of relations. (See Auger and Barrière 2008 for more examples.) Each of these was designed to meet the specific needs and goals of the situation in which it was to be used, and to integrate them into a given workflow or software suite. For example, depending on needs, these systems might attempt to identify the terms or other items linked by a relation marker to accelerate the process of analyzing the extracted KRCs. However, this is a complex task (cf. Section 2.2.2.1).

In a number of recent projects (e.g., León-Araúz, San Martín and Faber 2016; Weilgaard Christensen 2019; San Martín, Trekker and León-Araúz 2020), researchers have investigated the use of word sketches produced by the online corpus tool, Sketch Engine,⁵ as an alternative for manual querying or developing purpose-built software tools – which involve a considerable investment of time and effort (and expertise, in the case of custom tools). In addition to producing reusable pattern sets, this tool offers the potential of automatically extracted syntheses of the grammatical and collocational behaviour of search words in corpora. The approach involves the creation of CQL expressions (combining regular expressions and part-of-speech tags) which can then be saved in the form of a *sketch grammar*, and reapplied (in combination with the standard statistical analysis functions offered by the tool) to identify useful term pairs in context in text corpora. Designing these expressions, as in almost any development of a KP set, requires the iterative refinement of patterns to account for variations (cf. Section 3.2 below), and to optimize performance (cf. Section 4). Nevertheless, the potential for reuse of a sketch grammar to quickly produce an easily interpreted portrait of potential relation occurrences in a variety of corpora without the need of advanced programming skills is a promising direction for future work.

5. Sketch Engine. Lexical Computing (<https://www.sketchengine.eu/>) Accessed 23 December 2020.

2.2.2.1 *Some challenges of KP use for corpus analysis*

As Meyer (2001) explains, the nature of corpora is such that KRCs rarely appear as perfectly formed definitions ready for use in terminology resources. The high levels of variability in the form and expression of these contexts, as well as subtle and complex linguistic cues and background knowledge that are essential to human understanding (Meyer et al. 1999, 258; Aussenac-Gilles and Condamines 2012, 106), generally entail verification and interpretation by the user to ensure that the knowledge contained is properly identified and classified.

One reason that this human interpretation can be so critical is that many patterns may be ambiguous (Meyer 2001; Aussenac-Gilles and Condamines 2012). For example, the relation marker *are* may indicate a generic-specific relation (e.g., *multi-word terms are terms that...*), but may also indicate another type of conceptual characteristic, or no characteristic at all (e.g., *multi-word terms are challenging to delimit*). This is true to an even greater extent of many paralinguistic and grammatical patterns (hence their more restricted use). Taking into account the context and corpus in which a candidate KRC is found may in some cases help to disambiguate occurrences. Condamines (2002, 2008b) for example, observes that KPs may be more likely to indicate a specific relation in a particular subject field or text genre. The same may be said for KPs that combine with certain semantic classes or categories of terms (e.g., processes, entities) to indicate certain relations. Nevertheless, it is generally impossible to resolve all ambiguities without human intervention.

It is also challenging – even for a human, but certainly for automated processes – to identify appropriate terms participating in a specific relationship of interest. While the ideal of relation extraction might be the identification within a single sentence of two terms linked by an unambiguous marker of a direct relation between them (e.g., *a parallel corpus is a type of corpus*), the relationship as expressed in a corpus might be a more distant one (e.g., a term linked not to its immediate generic but to a broader one, e.g., *a parallel corpus is a type of text collection*). The perspective might vary with an author's purpose or situation (e.g., in the case of multidimensionality [Bowker 1993, this volume],⁶ e.g., *a parallel*

6. Multidimensionality (Bowker 1993) describes a situation in which concepts may be classified in a variety of different ways, depending on perspectives. For example, in the upper levels of a concept hierarchy reflecting generic-specific relations, vehicles may be categorized according to the environment in which they are used (e.g., surface vehicles, air vehicles, aquatic vehicles), how they are powered (e.g., motorized vehicles, self-propelled vehicles), or what they transport (e.g., cargo vehicle, passenger vehicle), with the remaining distinctions potentially appearing at lower levels of the hierarchy. It is not uncommon to find examples of varying hierarchical structures within the same corpus, making establishing a single unanimous conceptual structure challenging.

corpus is a type of translation resource). In addition, one or both of the terms might not appear in a standard form (e.g., be replaced by a variant, a different part of speech, or a pronoun in an anaphoric reference, or be modified by conjunction or disjunction) (Meyer 2001; Marshman 2007b; Aussenac-Gilles and Condamines 2012). Anaphoric reference is particularly common when knowledge-rich contexts are made up of more than one sentence, which adds additional challenges for their identification and extraction (Meyer 2001; Aussenac-Gilles and Condamines 2012). Non-standard order or interruption of the pattern structure, even within a single sentence, is also relatively common (Marshman 2007a; León-Araúz, San Martín and Faber 2016) (e.g., *a parallel corpus, a valuable tool for translation-oriented research, is a type of corpus that...*).

Even when relations and the items they link are correctly identified, it is necessary to take care when interpreting and representing KRCs, as certainty (or, more relevantly, uncertainty) may be indicated by items in the context, including hedges (e.g., *knowledge patterns sometimes identify relations*), modal verbs (e.g., *knowledge patterns can identify relations*), quantifiers (e.g., *some knowledge patterns identify relations*) and negation (e.g., *some knowledge patterns do not identify useful relations*), alone or in combination (e.g., Meyer 2001; Marshman 2008).

2.2.3 Use for corpus building

Even before the corpus analysis stage, Meyer also suggests (2001, 299) that knowledge patterns might be used to assist in the process of corpus building. She hypothesizes that since text types in which there is a large knowledge gap between author and reader are likely to contain more KRCs (an observation also made by Pearson 1998), knowledge patterns might be used to help spot particularly knowledge-rich text types in the corpus-building process, and potentially to help classify them.

This application has been investigated by Agbago and Barrière (2005) and Barrière (2006), who explore an iterative strategy for building a corpus semi-automatically, using a tool called TerminoWeb.⁷ This application used seed terms and a “starter” set of relatively domain-independent knowledge patterns for a variety of relations (e.g., generic-specific, part-whole, cause-effect, object-function, synonymy) to search for documents on the Web, and then rank them in decreasing order of knowledge pattern density, as an indication of their knowledge-richness. The user could then view the occurrences of the KPs in the texts to assist them in determining which texts to retain, and also potentially identify additional seed terms (manually or through term extraction), which could then be used for further iterations of searching for corpus texts. To help to adapt to different domains and/or

7. Unfortunately, this tool is no longer available.

text types of interest, users could also specify additional KPs to be added to the initial set. Despite its promise, the authors are careful to note that this semi-automatic approach is complementary to (rather than a replacement for) conventional – but less easily automated – corpus text selection criteria of domain, text type, quality, and reliability.

An additional application, as an alternative to building a corpus *per se*, involves searching directly on the Web (e.g., Halskov and Barrière 2008). The WWW2Rel tool was developed to help terminologists to expand or update ontologies by searching for snippets directly on the Web. It assists in term-oriented searching for useful occurrences of relations by allowing the user to identify a term of interest, and construct a query that combines that term with a set of KPs for the desired relation(s). It then ranks the results based on factors such as the frequency with which a candidate relation participant appears and the range and nature of KPs with which it combines, filters the list of potential relation appearances using statistical calculations and analysis of their forms (e.g., to group together terms that share the same head), and presents these to the user for consideration. The use of the Web as a corpus of course entails the challenges associated with the varied and uncontrolled nature of the content, including the potential to encounter dubious or patently incorrect information (Halskov and Barrière 2008, 25–26). However, the authors consider that this is an acceptable risk, at least for certain applications.

2.2.4 Other uses of KPs

With new perspectives on terminology opening up thanks to the availability of large text corpora and the potential for their analysis in various perspectives (cf. Faber, this volume; L'Homme, this volume), and with models of terminology resources in these perspectives evolving to accommodate more kinds of information, we have begun to see projects focussing on the use of KPs and KRCs not only behind the scenes of term banks or in the creation of ontologies, but also at the forefront of specialized language resources. Marshman, Gariépy and Harms (2012), for example, discussed the potential of bilingual KRCs annotated with relation information as a resource to help language professionals write and translate (and better express conceptual relationships) in specialized domains. In this section, we briefly describe two recent projects that have integrated KRCs containing KPs in specialized resources.

The CRISTAL project (Condamines et al. 2013; Josselin-Leray et al. 2014) was a collaborative project launched in late 2012 involving researchers from the Université de Nantes, Université Toulouse-Le-Mirail, and the University of Geneva, and an enterprise partner, Lingua et Machina. Inspired by the need to provide more and richer contexts to users of terminology resources (and particularly translators), the researchers set out to build a platform that would identify and display KRCs

extracted from comparable corpora. Their goal was to select contexts that would provide relevant co-occurrences and other information, including relation information, which might assist users in research and translation tasks. Their approach involved identifying KPs in English and French comparable corpora, extracting KRCs containing these patterns, and presenting them in an integrated interface for consultation during translation. They followed up this development with a mixed-methods approach to evaluating the usefulness of the information and its presentation (cf. Section 4 below).

The EcoLexicon (Faber, León-Araúz and Reimerink 2014; León-Araúz, San Martín and Faber 2016; Reimerink, León-Araúz and Faber 2016) is a multilingual TKB for the field of Environment Science that makes use of KPs for the discovery of relations, which can then be stored in a database and displayed in a dynamic, customizable graphical format that allows users to visualize the most relevant relations for the perspective and concept that interests them (cf. Faber, this volume). EcoLexicon makes use of KRCs not only behind the scenes, in the identification of relations to be represented in the database, but also in the form of *conceptual concordances*, i.e., text excerpts that show how conceptual relations appear in corpus texts. This customization of the display allows the user to circumvent some of the challenges of multidimensionality, and ensures that the richness of the information contained in the TKB does not become overwhelming or confusing (or too time-consuming to interpret). Users can thus target the most relevant information for their purposes. However, this does require the meticulous alignment of approaches and perspectives throughout the development and sections of the resource.

3. Preparing patterns for use

As previously noted, pattern-based applications typically involve the creation of sets of KPs for specific language(s) and relations, and then formally representing them so that corpus analysis tools can spot their occurrences and extract candidate KRCs for analysis. A variety of methods for discovering, translating, formalizing, and searching for patterns can be identified, each with its own advantages and challenges.

3.1 KP discovery and selection

The discovery of KPs often relies on an iterative process referred to as *bootstrapping* (e.g., Hearst 1992; Morin 1999; Condamines and Rebeyrolle 2001). This process begins with pairs of terms that are known to be linked by a given relation. A search for these terms in proximity to one another may be used to identify a candidate set

of relation markers. These in turn may be used to identify new pairs and expand the set of related terms, which can then be used to discover new markers, and so on. Other strategies include the use of dictionaries or thesauri to identify synonyms and variants to expand pattern sets (e.g., Ahmad and Fulford 1992; Barrière 2004). Ultimately, a set of relatively frequent and reliable markers can be established as candidates for inclusion in KP sets that are reusable for future applications (at least, within the limitations imposed by patterns' variability).

For bilingual or multilingual work, pattern sets may be identified independently in each language in comparable corpora. Alternatively, some projects (e.g., San Martín, Trekker and León-Araúz 2020) have used dictionaries, thesauri or other lexical resources to identify equivalents of markers that can then be integrated into KP sets. More rarely (e.g., Marshman, Gariépy and Harms 2012; San Martín, Trekker and León-Araúz 2020), parallel corpora may be used to identify translation equivalents of an initial set of markers in a source language. These may then be searched in the target language for back-translations to identify previously unidentified markers in the source language.

3.1.1 *Challenges of KP discovery and selection*

The wide range of potential relation markers and the various forms in which they may occur (both within and between domains) make developing exhaustive lists of knowledge patterns a practically impossible task (e.g., Condamines 2002).

Another commonly observed issue (e.g., Meyer 2001, 288; Condamines 2008b; Aussenac-Gilles and Condamines 2012) is that relations may differ in importance between domains, and this may have an impact on the usefulness of KPs. For example, cause-effect relations are expected to be particularly important in the Life Sciences. In addition, the range and classifications of relations (e.g., temporal versus physical parts in the part-whole relation, or different subtypes of causal relations [cf. Garcia 1997; Barrière 2001]) – and the associated KPs – may also differ with the situation and context of use, and the semantic classes of the terms they link (e.g., substances versus processes) (Meyer et al. 1999, 262; Condamines and Rebeyrolle 2001; Faber, León-Araúz and Reimerink 2014). Even when relations are relatively domain-independent, the frequency of occurrence of the KPs that indicate them may vary considerably in different domains and text types (e.g., Meyer et al. 1999; Jacques and Aussenac-Gilles 2006; Condamines 2008b; Marshman, L'Homme and Surtees 2008a). The combination of these factors makes selecting and structuring an optimal set of these patterns a time-consuming task that must often be repeated (at least to some extent) for each new project.

As a linguistically based approach, KPs are language-dependent and sets must be established for each language in which they are used. In bilingual and multilingual applications, differences in the nature and frequency of patterns, the form

of the items they link, and the nature and function of interruptions may affect the efficiency of pattern sets and forms, and require adaptation to produce similar results (Marshman 2007a). Independently identified bilingual or multilingual KP sets will tend to reflect usage in each language, but may not be entirely parallel in their number and content (e.g., Marshman 2007a) and their creation is more time- and resource-intensive, requiring fairly large corpora in each language. Conversely, sets that have been translated using dictionaries or other lexical resources will more likely be parallel in the languages and are quicker and less resource-intensive to produce. However, they may not reflect real language usage, and thus may not be as effective. Approaches using parallel corpora are somewhat more likely to reflect usage (although the target language is inevitably influenced to some extent by the source language) and are likely to be less time-consuming to produce than independent lists, though somewhat more so than dictionary look-ups. However, this approach evidently depends on access to a suitable bitext corpus in the desired subject field and text type.

3.2 KP representation

Once patterns are identified, they must be represented in a way that makes their re-use in semi-automatic applications feasible and efficient. As discussed in Section 2.2.2, in their simplest form, KPs may be implemented as character strings (sometimes accompanied by wildcard characters or other simple operators) that may be used – often in proximity to one or more terms represented in a similar way – by corpus analysis software to produce concordances that can then be screened manually. However, such a simple model requires the time-consuming and meticulous construction of the query to represent each KP, as well as careful interpretation of the results, which are likely to be quite noisy (cf. Section 4).

The refinement of KP representation often entails the use of part-of-speech (POS) tagged corpora and the representation of the patterns as a sequence of POS and regular expressions that restrict the nature and placement of the triple (term + marker + term) forming the pattern. For example, Hearst (1992) focuses on the creation of regular expressions that combine noun phrases (NP) occurring before and after lexical markers of the generic-specific relation, for example:

- NP *such as* {NP1, NP2 ..., (and|or)} NP_n
- NP {,} *including* {NP,* {or|and} NP
- NP {,} *especially* {NP,* {or|and} NP)

Condamines and Rebeyrolle (2001, 138–140) provide a more detailed look at the sequences in which candidate terms (CTs) might be found in their patterns for generic-specific and part-whole relations, for example:

def_det + CT1 + Vtobe (present) + undef_det + {*kind, type, etc. of*} CT2 +
 {relative clause, past participle, present participle, adjective}
 def_det + CT1 + Vtobe (present) + undef_det + CT2 + {relative clause,
 past participle, present participle, adjective}
 CT2 + {*contains, encompasses, includes, etc.*} + CT1

For certain relations, Condamines and Rebeyrolle (2001, 131) add additional constraints on the linked term, to disambiguate markers and further reduce noise. They consider, for example, noun/noun phrase and verb terms, and filter potential KRCs to retain only those in which the linked terms share a head (e.g., as in *knowledge patterns encompass lexical knowledge patterns, grammatical knowledge patterns, and paralinguistic knowledge patterns*). Analysis of the extracted KRCs also allows the authors (2001, 139–142) to identify classes of concepts (e.g., human, activity, time) expressed by the terms that might be used to further refine patterns to target specific types of occurrences.

3.2.1 *Challenges of KP representation*

The needs to simultaneously provide a fine-grained description of the sequence and the items that appear in it and still remain flexible enough to deal with the variations inherent in language (cf. Section 2.2.2.1) combine to make the task of KP representation a highly challenging one. Moreover, as previously discussed, relation types and sub-types and various levels of granularity may be required to reflect domain structures and the processes and concept classes necessary for domain description. The more fine-grained the description of the relations and their participants, the more complex the representation of the knowledge patterns required. KP representation has a direct effect on the performance of pattern-based approaches in general.

4. Evaluation of pattern-based approaches

Evaluation of KP-based knowledge-extraction approaches essentially gravitates around the most fundamental opposition in any information extraction task: that of precision and recall (Meyer et al. 1999; Malaisé, Zweigenbaum and Bachimont 2005; San Martín, Trekker and León-Araúz 2020). Essentially, precision evaluates how many of the results produced using a KP or a KP-based approach or tool are

in fact useful (and how many constitute *noise*), whereas recall would ideally measure how many of the potentially useful results available are located. Of course – at least without an extensive and very time-consuming manual corpus analysis – accurately evaluating recall is an almost impossible task on any meaningful scale. Moreover, recall of KPs is commonly considered to be very low in general (Auger and Barrière 2008), leaving precision as the most commonly used indicator, sometimes complemented by the raw number of occurrences identified or by other more indirect measures.

In some cases, evaluation may call on a comparison with a gold standard lexical resource (e.g., the Unified Medical Language System [UMLS] for Halskov and Barrière 2008) either to validate the information found or to identify novel contributions that may be located using KPs, with resulting improvements in the resource produced (e.g., as suggested by Aussenac-Gilles and Jacques 2008). In one project, Halskov and Barrière (2008) look at the association of KPs with two sets of known term pairs, one set linked by the desired relation, and the other linked by another type of relation. Comparing the occurrences of the KPs with each of the sets of terms allows the researchers to estimate the precision of the patterns for identifying the relation in question (assuming that more precise terms would occur more often with terms linked by the desired relation, and less often with the others). Alternatively, comparisons may be made with other approaches, such as those based on statistical or machine-learning techniques (e.g., Kamel et al. [2017], which uses pattern-based approaches as a baseline to evaluate the performance of a machine-learning approach to relation extraction). More commonly, simple human evaluation of the presence or absence of the desired relation in extracted KRCs has been used.

Challenges are inevitable in the evaluation process. When comparing results to existing resources, there will be issues surrounding the coverage – and potentially the perspective – of the resource in question, which may affect results. For human evaluations, the subjectivity of the evaluation may be an issue (though one that may be mitigated by using multiple evaluators and measuring inter-evaluator agreement [cf. Halskov and Barrière 2008; Kamel et al. 2017]). An even more difficult challenge to overcome is the time- and labour-intensive nature of human evaluation.

At the level of individual patterns, as with any query, refining forms to eliminate as much noise as possible (i.e., to have the highest possible precision) entails a drop in recall (i.e., the loss of some potentially relevant results). This is a particularly sticky issue for KPs, which often produce what Meyer et al. called (1999, 265) *good noise*, namely, contexts that assist in the acquisition of domain knowledge, even if they do not contain the anticipated relation and terms. In fact, differentiating between “hits” and good noise can be immensely challenging, given variability of KPs (cf. Section 2.2.2.1). Halskov and Barrière (2008, 24–25) recall that in addition

to undeniably useful contexts in which KPs link two domain terms that are clearly linked by a desired relation, results might also illustrate a relation other than the one intended (as is common in cases of KP ambiguity). Furthermore, they might indicate the desired relation but not between terminologically relevant items, or include some expression of uncertainty that limits the potential applications of the information. In more automated applications (e.g., automatic construction of ontologies), it may be necessary for a pattern to return a very high proportion of results that contain the prototypical two terms linked by the lexical relation marker with minimal interruption. However, in applications that are semi-automatic or primarily manual, human users may well be able to make good use of imperfect results by expanding the scope of their reading to deal with intersentential KRCs and anaphoric references. They would thus be able to infer or interpret the results to deal with broad or imprecise terms or terms that vary from the prototypical noun form, or KRCs that convey some level of uncertainty (Meyer et al. 1999; Meyer 2001).

The issue of multidimensionality may also contribute to difficulties in evaluation, as it can entail challenges even when relations are obviously present and link two clearly identifiable terms, since the specific relation identified may not align with the researcher's perspective or goals.

Another issue in the evaluation of pattern-based applications is the variation in performance of individual KPs and even of entire sets of KPs for specific relations, depending on the corpus (including subject field and text type or genre, cf. Section 3.1) (Barrière 2004; Aussenac-Gilles and Condamines 2012). Even when an accurate measure of the performance of a KP set in a given corpus can be achieved, there is little guarantee that this performance will necessarily be reproducible in a new corpus. Efforts may be made to control (or at least to identify) domain- and text-type-related variation (e.g., Marshman, L'Homme and Surtees 2008a; San Martín, Trekker and León-Araúz 2020), but the limitations of such studies are still difficult to identify. For example, San Martín, Trekker and León-Araúz (2020, 5962) observed lower levels of performance in some corpora (e.g., of legal texts) despite efforts to identify and refine patterns in a variety of domains and text types.

While it is acknowledged that KPs will often link terms belonging to particular semantic classes or categories, it is nevertheless not helpful if they are tied too specifically to individual terms. Testing of the "term range" of KPs by investigating the variety of terms with which KPs combine in a given sample (Halskov and Barrière 2008) can help to evaluate how broadly applicable KPs are likely to be and help to eliminate those that are unlikely to be useful for finding new information.

Recognizing the challenges of evaluating performance of individual KPs for extracting KRCs, Barrière (2004) proposes a measure of relative productivity by attempting to determine whether a given pattern identified a higher-than-expected proportion of KRCs, in comparison with others in the KP set. While this does not

allow for an overall prediction of productivity for a KP set, it may allow for more ambiguous or problematic patterns to be identified, and either refined or discarded.

Evaluating other types of KP-based applications may entail even greater challenges and require creative strategies and comparisons, as well as considerable effort.

For corpus-building, in the TerminoWeb project (Agbago and Barrière 2005; Barrière 2006), a comparison with simple search engine searching and the use of the first hits retrieved shows considerable advantages in KRC density with TerminoWeb. The results approximate those obtained from corpora built manually by terminologists (with a considerable investment of time and effort) in the fields of composting and scuba diving.

In the CRISTAL project, the usefulness of the KRCs identified with KPs for student and professional translators is evaluated via a rich mixed-methods approach that includes screen recording, keystroke logging, questionnaires, and interviews (Josselin-Leray et al. 2014; Picton and Josselin-Leray 2019). This rich data provides a wealth of opportunities to explore the behaviour of users and their interactions with the KRCs, as well as to learn about how they judge the usefulness of KRCs containing KPs. However, the enormous amounts of data collected in such a mixed-methods approach, and the need to triangulate them to draw conclusions, entail an intensive and lengthy analysis (Picton and Josselin-Leray 2019). Evaluation of the CRISTAL project has produced promising feedback, including preferences for KRCs compared to other types of context (Josselin-Leray et al. 2014). However, Picton and Josselin-Leray (2019, 302–305) note that the exact definition of *richness* in such applications needs to be clarified, as the usefulness of contexts may vary considerably with the user's purpose and personal requirements.

5. Future directions

As the tools and resources available for exploiting knowledge patterns and representing the product of their use evolve, so do the research directions in this area.

The availability of software tools such as Sketch Engine, which allow users to combine standard statistical analyses with fairly easily programmed (or adjusted), customized and reusable queries, may open the door to more widespread use of KPs both by researchers and by more 'casual' users (e.g., terminographers or even translators) who would otherwise not have the time or capacity to develop a customized tool, but who may be able to make use of pattern-based search strategies. This use by professionals would fulfill one of the initial goals of some of the early developers of pattern-based approaches that has – due in part to various technical

limitations and the complexity of domain-independent KP set development – not as yet become a reality. Nevertheless, this possibility is coming closer with the growing range of available KPs and information about them, combined with the potential to more easily share KP sets, although the issues of variability are still relevant. The potential contribution to research offered by the option to test KP sets in a wide range of corpora for diverse applications is also considerable.

As in many computer-assisted tasks today, there is interest in the role of Artificial Intelligence and machine learning in future approaches to relation identification and extraction. Developments in machine-learning techniques will almost certainly open new avenues as alternatives or complements to pattern-based approaches. The potential for combining various approaches to benefit from their different strengths is promising.

Given the established usefulness of corpora for both conceptual and linguistic research in terminology, translation, and other language activities – but also the time-consuming and labour-intensive nature of corpus-building – it is surprising that the use of KPs to help build corpora has not received more attention. Further research in this area would be welcome. A particularly glaring gap is that of corpus use among professional translators, which to date does not appear to reflect the potential of these resources for both contextual and linguistic research. More adapted tools for both corpus construction and querying could help to encourage corpus use by some professionals. Similarly, the inclusion of KRCs containing KPs in specialized language resources for translators, writers, and other language professionals – as well as trainees in these professions – is an area ripe for expansion. Some initial projects in this area, and the generally positive response to the integration of KRCs in terminology resources, suggest that development in this direction will be valuable to language professionals and students, as well as researchers.

Terminology and distributional analysis of corpora

Ann Bertels

University of Leuven

This chapter discusses the theoretical and methodological principles of distributional semantic analysis. Over the last decade, Distributional Semantics has become very popular in Corpus Linguistics, building on very large corpora and extracting useful semantic information for numerous applications. In our “big data” era, Artificial Intelligence (AI) is carving its way into our daily life. AI’s algorithms in Natural Language Processing (NLP) learn from text collections with easily accessible information in order to find and predict knowledge patterns. This chapter explores the use of distributional analysis for terminological needs, i.e., in specialized domains. It focuses on what distributional analysis stands for, how it works, how it can be used for LSP and Terminology, and why it is useful for terminological needs.

Keywords: Distributional semantics, vector space models, specialized corpora, semantic similarity, semantic relatedness, co-occurrence analysis

1. Introduction

Within the framework of Corpus-based Terminology, the distributional analysis of specialized corpora is mainly a question of distributional semantic analysis. It takes into account the usage context or linguistic environment of a term in specialized corpora in order to gather semantic information on the term (i.e., its co-occurrences).

Generally, this field of study and research is referred to as Distributional Semantics (DS), because the distributional information for a word (or term) is used to compute semantic similarity or semantic relatedness. The underlying idea of DS is that words or terms with similar distributions have similar meanings.

Since Corpus-based Terminology (see Marshman, this volume), DS and distributional analysis have also yielded very interesting results (L’Homme 2020c)

in a wide range of specialized domains and for various languages, this chapter provides an overview of recent research in DS from a terminological perspective. Terminology can certainly benefit from the use of distributional methods, because the analysis of word co-occurrences in specialized corpora provides useful information for terminology extraction and domain-specific vocabulary expansion, ontology building, extraction of semantic relations, detection of paradigmatically related terms, and detection of terminological variation. The rest of the chapter is organized as follows. Section 2 explains the general theoretical and methodological principles of Distributional Semantics as well as the different type-level and token-level DS models, with window-based and syntactic co-occurrences in both count-based models and prediction-based models. In the context of this theoretical and methodological background, Section 3 explains all the terms, concepts, and underlying principles required for a good understanding of the distributional analysis of specialized corpora. This is followed in Section 4 by a discussion of challenges and solutions, in this area of study. Finally, Section 5 lists the conclusions that can be derived from this study and points the way to new research paths and opportunities in distributional semantics.

2. Distributional Semantics

2.1 Theoretical and methodological principles

Distributional semantic analysis or Distributional Semantics (DS) builds on the theoretical foundation of the “Distributional Hypothesis” (Harris 1954, 1968), which states that words with similar meanings tend to occur in similar linguistic contexts. This ties in with Firth’s famous adage “You shall know a word by the company it keeps” (Firth 1957, 11). Although Firth is referring to collocation analysis, the idea is the same: co-occurrences of a word provide access to its meaning. Words with similar meanings appear in similar contexts and have similar co-occurrences, whereas words with different meanings have (very) different co-occurrences. For the benefit of readers who are not familiar with DS and distributional analysis, we will use a simple example including the words *car*, *bike* and *doctor* throughout this section in order to illustrate the main concepts of DS. The words *car* and *bike* appear in more similar contexts and will thus share more co-occurrences (e.g., *traffic*, *road*, ...) than *car* and *doctor*. The word *doctor* mostly appears in very different contexts (e.g., *hospital*, *nurse*, ...) with respect to *car*, and, as a consequence, will share only very few co-occurrences with the word *car*.

Distributional information or co-occurrence information can be extracted by using large text corpora. Accordingly, DS is a bottom-up approach in that a

word's meaning emerges from its distributional properties in large text corpora without the use of other knowledge sources (Fabre and Lenci 2015). Distributional properties can be statistical or syntactic co-occurrences; and semantic analysis can be performed at either the type- or token-level, which gives rise to various models. Distributional Semantic Models (DSMs) also vary with respect to the way co-occurrence information is stored and the way distributional representations are built (Heylen and Bertels 2016; Lenci 2018) (See Section 2.2 for an overview of commonly used DSMs.) Most of these models follow the same four-step procedure (Baroni and Lenci 2010; Turney and Pantel 2010; Fabre and Lenci 2015; Fabre 2015; Clark 2015; Lenci 2018) though with varying parameters.

– Step 1: Extraction of co-occurrences

Co-occurrences of a word or term (as a node) and its linguistic contexts (as collocates) are extracted from a large text corpus, and are counted and stored in a co-occurrence vector.¹ For example, the co-occurrence vector of *car* will store the collocates *traffic* and *road* as well as the number of times they occur with *car*, i.e., the co-occurrence frequency or co-frequency of *traffic* and *road* with *car*. Next, co-occurrence vectors of two or more words are stored in a co-occurrence matrix, with the words as rows (e.g., n rows), with their linguistic contexts or collocates as columns (e.g., m columns) and with their co-occurrence frequencies as entries in the $n \times m$ co-occurrence matrix. In the co-occurrence matrix of our example, the words *car*, *bike* and *doctor* are used as rows, while their collocates *traffic*, *road*, *hospital* and *nurse* are used as columns. The co-occurrence frequency or co-frequency of *car* and *traffic* is stored as an entry in the row *car* and in the column *traffic*.

– Step 2: Weighting

Most models, however, do not use raw co-occurrence frequencies, but rather weighted co-occurrence information, for example, the collocational strength between node and collocates (e.g., between *car* and *traffic*) in order to take

1. A vector is a mathematical concept. It conveys information about an object and can be used to represent the object in a two-dimensional, three-dimensional or high-dimensional space, by means of an arrow with a magnitude (length) and a direction. Vectors are commonly used in Distributional Semantics in order to carry information about the neighbouring words of a word or term (node). This means that vectors convey distributional information about words or terms. As a consequence, distributional similarity is defined by means of vector similarity. Words or terms with similar neighbours (or co-occurents) are represented by similar vectors, which have a similar magnitude and a similar direction in vector space. These similar vectors can be visualized as arrows with similar length (their magnitude) and similar direction. The more similar the direction of the vectors, the smaller the angle between the arrows in 2D, and thus the more similar the words or terms.

into account the relevance or the statistical significance of the linguistic context. Association measures have been explored and compared in various studies (Evert 2004; Pecina 2010; Evert et al. 2017; Garcia, García-Salido and Ramos 2019). In a wide range of studies and tasks, Positive Pointwise Mutual Information (PPMI) was found to yield the best results (Turney and Pantel 2010; Fabre and Lenci 2015). Nevertheless, as weighted co-occurrence information is extracted from large text corpora and most of the words co-occur with only a few collocates, most of the entries in the matrix are empty, giving rise to a rather sparse co-occurrence matrix. This problem is known as data sparseness. For example, the words *car* and *bike* do not co-occur with *hospital* and *nurse*, and, consequently, the rows, *car* and *bike*, in the co-occurrence matrix of our example will be empty in the columns, *hospital* and *nurse*.

– Step 3: Dimension reduction

To solve the data sparseness problem, the information in the co-occurrence matrix (with m dimensions) is densified using dimension reduction techniques, such as Singular Value Decomposition (SVD). In this technique, the original sparse m -dimensional weighted co-occurrence matrix is mapped onto a low-dimensional and thus dense matrix. An alternative to dimension reduction techniques is Random Indexing (Sahlgren 2005, 2006), which builds a low-dimensional representation by assigning each word to a random vector by means of the vectors of the co-occurring words (Fabre and Lenci 2015).

– Step 4: Calculation of similarity

Semantic similarity between words can be computed by means of similarity measures. Of the measures compared in NLP, the cosine angle was found to be the best performing distance measure for observations represented as vectors (Padó and Lapata 2007; Sahlgren 2008; Fabre and Lenci 2015). Cosine angle determines the (semantic) similarity between two words by calculating the angle between their vectors. Words with similar vectors cluster together in multidimensional space, because the angle between them is small. In our example, the words *car* and *bike* will cluster together or be plotted very close to each other, and will thus be considered semantically similar, because the angle between their vectors is very small, compared to *doctor*, which will appear closer to *surgeon* or *patient* than to *car*. Semantic similarity between words can be mapped onto spatial proximity multidimensional space, which can be visualized on a 2D-plot or a 3D-plot (Bertels and Speelman 2016).

2.2 Overview of Distributional Semantic Models (DSMs)

2.2.1 *Differences with respect to the linguistic context*

In step 1, co-occurrences of a word or term (as a node) and its linguistic contexts (as collocates) are extracted from a large text corpus (see Section 2.1). Collocates can be: (1) window-based; (2) syntax-based; or (3) document-based. Lenci (2018) gives a comprehensive overview of the different types of linguistic context with examples. For a comparison between window-based and syntax-based models, see Tanguy, Sajous and Hathout (2015).

1. Window-based or word-based models look for co-occurrences of a word or term within a context window of usually five words to its left and right. Some models distinguish between left and right context, i.e., directed window-based collocates, while others do not make this distinction, i.e., undirected window-based collocates. Window-based models are also referred to as bag-of-words (BOW) models, because they do not take syntactic dependencies into account (Heylen, Speelman and Geeraerts 2012; Ferret 2014; Bernier-Colborne 2014a; Bertels and Speelman 2013, 2014a, 2014b; Périnet and Hamon 2014; Azoulay 2017; Rojas-García and Faber 2019). Since they are fast and easy, window-based models are typically used to identify the syntagmatic relations between words. Similar words tend to share the same neighbours.
2. Syntax-based models define co-occurrences of a word or term on the basis of syntactic dependencies and thus require a parsed corpus (Morlane-Hondère 2013; Morardo and Villemonte de La Clergerie 2013; Fabre et al. 2014; Fabre 2015; Brunet, Ferret and Tanguy 2020). Syntactic dependencies can be used only for the identification of syntactic collocates (i.e., dependency-filtered syntactic collocates), or can also be included in the context information for further exploitation in the next steps of the distributional analysis, in the case of dependency-type syntactic collocations. Syntax-based models are most relevant for the identification of paradigmatic relations and taxonomically related words. Similar words share the same dependency relations.
3. Document-based models consider the whole text for the extraction of co-occurrences of a word or term and are typically used in Latent Semantic Analysis (LSA) (Landauer and Dumais 1997), where they provide information on associative relations. Similar words appear in the same document.

Multimodal Distributional Semantic Models go beyond linguistic context: they include visual context, such as images, building on recent advances in computer vision techniques (Bruni, Tran and Baroni 2014; Baroni 2016), and even natural sounds (Cassani and Lopopolo 2015; Lopopolo and van Miltenburg 2015). Vector-based

representations of visual and acoustic data and vector-based representations of textual data are integrated, thus producing a hybrid model of meaning. For example, Rotaru and Vigliocco (2019) integrate linguistic, visual, and affective data in a multimodal distributional model.

2.2.2 *Differences with respect to the level of analysis*

When identifying the co-occurrences of a word or term (as a node) in a large text corpus, most DSMs consider the node at type-level (Navigli 2012). Co-occurrence frequencies are then averaged over all the occurrences of the node in the text corpus and stored in a co-occurrence vector (Section 2.1, step 1). The resulting co-occurrence vector contains the semantic information conveyed by all the relevant co-occurrences of the node, which may be quite general. A DSM at type-level is also referred to as a type-level or a type-based Vector Space Model (VSM). A type-based analysis is typically used to distinguish between words, for example, for synonym extraction (Heylen et al. 2015).

For a more fine-grained semantic analysis, a DSM at token-level (Dinu, Thater and Laue 2012; Heylen, Speelman and Geeraerts 2012) considers the node at token-level and thus captures the semantic information of each individual occurrence of the node. After including the identification of the co-occurrences of the node, this DSM then retrieves the specific context of the first-order co-occurrences of the node, which are the second-order co-occurrences (Schütze 1998; Lemaire and Denhière 2006). Second-order co-occurrences (at type-level) give access to the semantics of the first-order co-occurrences of the node and, by extension, to the individual occurrences of the node. The resulting co-occurrence vector of a token-level DSM is less sparse than the averaged co-occurrence vector of a type-level DSM. Rather than a list of co-frequencies, it is a list of second-order co-occurrence vectors. By considering second-order co-occurrence information, a token-level DSM or token-based VSM captures differences in meaning between the uses of a single word. This is relevant to the following: (i) the analysis of polysemous words (Heylen et al. 2015) or polysemous terms (Bertels and Speelman 2014a); (ii) the study of polysemy issues in lexical lectometry (De Pascale 2019); and (iii) linguistic theory-building (Hilpert and Correia Saavedra 2017). A detailed explanation of type-based versus token-based semantic analysis is provided in Heylen et al. (2015) and De Pascale (2019).

2.2.3 *Differences with respect to the distributional representation*

The last distinction between DSMs involves the way the context information is stored. Distributional representation can be either count-based or prediction-based (Lenci 2018). In this regard, most DSMs are count-based models. This means that co-occurrences of a node are counted, usually weighted with respect to collocational

strength (Section 2.1, step 2) and stored in a vector. Hence, count-based models are also referred to as vector-based models.

Each dimension of a co-occurrence vector refers to a specific linguistic context of a node. Consequently, a co-occurrence vector contains an explicit representation of the context features of a node, and is typically very large but sparse. Co-occurrence vectors of two or more nodes are stored in a co-occurrence matrix, resulting in a high-dimensional but very sparse co-occurrence matrix. The very large number of context features is then reduced to a limited number of semantic dimensions for efficient computational processing.

Various dimension reduction techniques can be used to map the original high-dimensional (and explicit) sparse co-occurrence matrix into a low-dimensional (but implicit or latent) dense semantic space (Section 2.1, step 3), such as the following: (i) Singular Value Decomposition (SVD) in Latent Semantic Analysis (LSA) (Landauer and Dumais 1997); (ii) Principal Components Analysis (PCA); (iii) Nonnegative Matrix Factorization (NMF); and (iv) MultiDimensional Scaling (MDS). An alternative technique is Random Indexing (RI) (Sahlgren 2006).

Global Vectors (GloVe) (Pennington, Socher and Manning 2014) is a popular distributional model, based on weighted least-squares, which combines global matrix factorization and local context window methods. However, the interpretation of the resulting low-dimensional semantic space is not straightforward. Whereas the high-dimensional co-occurrence vectors and matrices refer explicitly to context features, it is not easy to assign a clear semantic value or interpretation to the latent dimensions (Lenci 2018). Consequently, the results of count-based DSMs may raise some questions on the kind of semantic information yielded. An important issue is that the semantic information quantified is not always compatible with what is traditionally regarded as semantic information. Count-based models are unsupervised methods because they identify clusters or patterns in unseen data.

Recent DSMs are prediction-based (Mikolov, Chen, et al. 2013), and inspired by methods of machine learning. Prediction-based DSMs rely on neural network algorithms. By learning to optimally predict which words occur together, they directly create a low-dimensional and dense (implicit) distributional representation of that word (Fabre and Lenci 2015; Lenci 2018). These low-dimensional representations are called *neural embeddings* or *word embeddings*² (Levy and Goldberg 2014).

2. *Word embedding* is a term used in NLP (Natural Language Processing) for the representation of a word by means of a vector containing distributional information about the word. The word can thus be represented as a very high-dimensional but very sparse vector. *Embedding* refers to the mathematical embedding from high-dimensional space into a denser vector-space with a lower dimension, by means of training methods used in neural-network language modeling.

Models based on neural networks are mostly supervised methods, since they learn from examples. They use training data before proceeding to the actual test data.

The word embedding toolkit Word2Vec (Mikolov, Sutskever, et al. 2013; Word2Vec 2013) is a shallow neural network with two layers. It contains two popular word embedding models: CBOW and skip-gram. CBOW (continuous bag of words) predicts a word based on the surrounding context words, is faster, and works well on large corpora. Skip-gram (skip-gram with negative sampling or SGNS) predicts the surrounding context on the basis of a word and is more reliable for smaller corpora. Word embeddings are very often used in NLP applications, such as Named Entity Recognition (NER) (Nadeau and Sekine 2007), because of their computational efficiency and scalability (Amjadian et al. 2018). For the creation of word embeddings on custom corpora, Python offers the GenSim library for researchers in NLP and information retrieval. GenSim (2020) can be used for topic modelling and similarity retrieval.

There have been various comparative studies on count-based and prediction-based models. Some found that prediction-based DSMs outperformed count-based DSMs (Baroni, Dinu and Kruszewski 2014; Levy, Goldberg and Dagan 2015), whereas others found that count-based matrix models yielded better results, especially on smaller datasets (Sahlgren and Lenci 2016). Overall, there are no significant differences (Fabre and Lenci 2015).

The latest distributional models are referred to as 'deep learning' models, because they take the context into account and use several intermediate hidden layers in the neural network. A well-performing model for word representation is ELMo (Embeddings from Language Models) (Peters et al. 2018), which relies on deep contextualized word representation. BERT (Bidirectional Encoder Representations from Transformers) (Devlin et al. 2019) is designed to pre-train deep bidirectional representations by using both left and right context in all layers. Since BERT is pre-trained on very large general corpora, it is useful for LGP analysis.

Recently, BioBERT (Bidirectional Encoder Representations from Transformers for Biomedical Text Mining) (Lee et al. 2019) was pre-trained on large biomedical corpora and can be used for analysis in the biomedical domain. A recent study compares BERT, ELMo, and GPT-2 Embeddings (Ethayarajh 2019), based on experiments that assessed BERT's performance (Mickus et al. 2020). For biomedical English, this is feasible because of the availability of large specialized corpora (based on articles in PubMed). Unfortunately, for other languages and/or other specialized domains, word embeddings are generated on small(er) specialized corpora (Tanguy, Fabre and Bard 2020).

3. Distributional analysis of specialized corpora

In the last five or six years, studies on general language corpora have mainly focused on improving the performance of DSMs by adjusting parameter settings (Lapesa and Evert 2014; Kiela and Clark 2014) and on deep learning techniques (Botov et al. 2019; El Boukkouri et al. 2019; Mickus et al. 2020). Because of smaller corpus-sizes in specialized domains, there has been less research on distributional analysis in specialized corpora than in general language corpora (Périnet and Hamon 2014). However, little by little, distributional analysis is being transposed to specialized corpora in various specialized domains and for various languages. Given the availability of English biomedical data (e.g., PubMed), distributional analysis on biomedical corpora in English will soon be quite productive (see Cohen and Widdows 2009 for an overview).

Meanwhile, distributional analysis has been used for various terminological applications, such as terminology extraction, taxonomy extraction, ontology building, semantic analysis, and terminological variation. This section gives an overview of recent research on distributional (semantic) analysis in specialized corpora from a terminological point of view.

3.1 Comparative studies applied to specialized corpora

Several comparative studies of large general language corpora have focused on the impact of window-based co-occurrences versus syntax-based co-occurrences as well as on the performance and efficiency of DSMs (Padó and Lapata 2007; Lapesa and Evert 2014; Kiela and Clark 2014). These comparative studies, however, did not lead to convergent results.

Following the example of these comparative studies on general language corpora, a very large set of experiments has been carried out on a small and specialized corpus in the domain of NLP in French (Tanguy, Sajous and Hathout 2015). On this tagged and parsed corpus of 2 million tokens, more than 2500 distributional models were compared, with varying parameters on several levels of the four-step procedure described in Section 2.1. These DSMs not only took into account the difference between window-based and syntax-based co-occurrences on the level of co-occurrence extraction, but also various association measures and window spans. In the specialized domain of NLP in French, the experiments showed that syntax-based models outperformed window-based models if the syntactic information was well processed. This might be due to the small size of the specialized corpus, because less informative models (e.g., window-based bag-of-word models) require more information in order to capture the semantic information (Tanguy, Sajous and Hathout 2015).

Within the shared task SemDis2014 (*Sémantique Distributionnelle*), the same small specialized corpus in the NLP domain was used for an exploratory comparative study of syntax-based DSMs with a view to optimizing distributional calculations (Fabre et al. 2014). A large set of parameters on different levels of the four-step procedure showed that, depending on the word class of the target word (i.e., the node), a better performance was achieved by fine-tuning the number, nature, and quality of the contexts in the DSM.

A recent study focused on the impact of document structure on the performance of DSMs in the NLP domain (Tanguy, Fabre and Bard 2020). Building on a corpus of research papers, different DSMs were compared, when sections of the document structure were omitted (section header, abstract, introduction, conclusion, etc.). Depending on the target word and the nature of the section removed, different results were obtained. However, the most interesting insight was that the most informative parts of the research papers had a lesser impact on the performance of the DSM than the less informative parts.

The Harbsafe-162 domain-specific data set evaluated DSMs in the electrotechnical domain. This study was carried out as part of the Harbsafe project of the German Commission for Electrical, Electronic & Information Technologies of DIN and VDE (Arndt and Schnäpp 2020). Distributional analysis was applied to terminological entries, consisting of one or more terms, term phrases, and a definition. The objective was to implement a sentence embedding model in this domain, as a first step towards the harmonization of terminologies of standards.

Within the specialized domain of environmental science, the EcoLexicon project, a terminological knowledge base on environmental science, focused on the geographic contextualization of data (Rojas-García and Faber 2019; EcoLexicon 2019). Several DSMs, with varying parameters, were applied to a small and specialized corpus of 7 million tokens in order to extract terms related to named rivers. The results showed that count-based models (with the log-likelihood association measure) performed better than prediction-based models. These results were in line with previous findings that count-based models outperform prediction-based models on small-sized corpora of less than 10 million tokens (Sahlgren and Lenci 2016).

3.2 Terminology extraction

The distributional analysis of specialized corpora has been very productive for terminology extraction and domain-specific vocabulary expansion in the biomedical domain. This is especially true for English, because of the availability of English biomedical data (e.g., PubMed). Recently, DSMs have also been used for terminology extraction for other languages than English and in other specialized domains.

In 2009, a distributional approach was used for terminology mining and terminology extraction in the specialized domain of social media, characterized by highly variable and dynamic word usage (Sahlgren and Karlgren 2009). The problem of continuously evolving vocabularies was addressed with the technique of Random Indexing (Sahlgren 2005, 2006), which achieved high recall and yielded promising results for capturing the shifting usage of words and terms in the specialized domain. However, this study revealed some pending issues related to the difference between semantic similarity and relatedness and the distinction between synonyms and antonyms. Random Indexing has also been used for the extraction of Japanese medical terminology in a patient blog corpus (Ahltorp et al. 2014, 2016). The experiments showed that when parameter settings were carefully chosen, distributional analysis could be used to successfully expand Japanese medical vocabularies, as has been done in English and other Germanic languages. In the domain of psychiatry in Spanish, distributional analysis has been applied to terminology extraction, based on the assumption that domain-specific terms, which refer to specialized concepts, show characteristic co-occurrence patterns (Nazar 2016).

In recent years, prediction-based models based on word embeddings have been used for terminology extraction in specialized corpora (Amjadian et al. 2018; Fan et al. 2019; Botov et al. 2019). The first study by Amjadian et al. (2018), in the mathematics domain, used both the GloVe model and the Word2Vec models, CBOW and skip-gram, in order to concatenate global vectors and local vectors into a single specificity vector. The word's distribution in both the general and specialized domain of mathematics provided the information necessary for its specificity to the specialized corpus and, hence, for its technical usage (Amjadian et al. 2018).

In the second study by Fan et al. (2019), word embeddings were applied to a corpus of electronic health records (EHRs) as a way to extract the terminology of dietary supplements. The Word2Vec model outperformed the GloVe model in the detection of semantically similar terms and in the retrieval of the semantic variants in clinical records (Fan et al. 2019). A study in the specialized domain of the labour market (Botov et al. 2019) applied the Word2Vec model and other DSMs to terminology extraction in order to match terms and concepts in job descriptions on online recruitment platforms as compared to those in professional standards.

3.3 Ontology building and taxonomy extraction

Distributional analysis has also been used for the (semi-)automatic creation of distributional thesauri and ontology building. For a specific entry, distributional thesauri list semantically related words based on the assumption that they share a similar distribution with the entry. Concepts and tools of Information Retrieval

and Distributional Semantics, including Word2Vec models, have successfully been implemented in research of large general language corpora for the construction of ontologies and thesauri (Claveau and Kijak 2015).

In a domain-specific corpus of socio-political documents, Distributional Semantics and the use of DSMs enabled the creation of an extended ontology, building on semantic similarities and differences for both single-word and multi-word lexical items (Kozerenko et al. 2019).

3.4 Semantic relations

Semantic or conceptual relations can be extracted from specialized corpora using various approaches, including Distributional Semantics and Knowledge Patterns (see Section 3.4 and see Marshman, this volume, for a detailed explanation of Knowledge Patterns).

Within the shared task SemDis2014 (*Sémantique Distributionnelle*) (see Section 3.1.1), Bertels and Speelman (2014a) and Bernier-Colborne (2014a) relied on Distributional Semantics for the extraction of semantic relations from specialized corpora. Bertels and Speelman (2014a–b) clustered first-order co-occurrences of target words with respect to shared second-order and third-order co-occurrences and then visualized them on a 3D-plot, in order to explore the semantic proximity between the collocates of each target word and thus obtain a better understanding of the polysemy of the target words. Similar experiments were carried out on French specialized corpora, both in NLP (Bertels and Speelman 2014a) and in a technical domain (Bertels and Speelman 2014b). Further fine-tuning of this methodological approach involved visualization by means of spinning 3D-plots (Bertels and Speelman 2016). Various experiments with varying parameters have shown that basic word co-occurrence information in a HAL-model can be used for the detection of the paradigmatically related terms of a target term, such as near-synonyms, antonyms, and hyponyms, both in NLP and in the environmental domain in French (Bernier-Colborne 2014a).

In a larger specialized corpus in English (50 million tokens) in the environmental domain, co-occurrence information was encoded in a three-way tensor, which led to more robust word space models for the extraction of semantic relations (Bernier-Colborne 2014b).

Building on Distributional Semantics and using Word2Vec models on a Wikipedia corpus in French, Wauquier, Fabre and Hathout (2020) carried out a comparative study of nominalizations. Their study refined the semantic distinctions between action nouns on *-age* and *-ion* in order to determine the level of technicality of these action nouns.

3.5 Knowledge patterns and semantic frames

Further analysis of semantic relations involves the use of knowledge patterns. Knowledge patterns are defined as “markers that convey semantic relations” (Rojas-García and Cabezas-García 2019, 121) and are “frequently used to extract conceptual information from a corpus” (121).

Distributional analysis has been combined with knowledge patterns for an in-depth analysis of the semantic relations between two similar terms belonging to the same cluster, in a corpus of 7 million tokens of English environmental texts (Rojas-García and Cabezas-García 2019), within the EcoLexicon project (see Section 3.1). In this study, a count-based DSM was used in order to assess the semantic similarity between terms and to generate semantic vectors, which were clustered automatically. The semantic coherence of these clusters was evaluated based on knowledge patterns in Sketch Engine (Kilgarrieff et al. 2014). This allowed for further exploration of the semantic relations between terms that clustered together, and eventually for the fine-tuning of semantic categorization.

In addition, distributional analysis carried out on specialized corpora of various sizes in English has been used for the extraction of specialized knowledge and the creation of semantic frames (Azoulay 2017). Semantic frames are conceptual structures which facilitate the interpretation of lexical meaning. In Frame Semantics, a frame is defined as “any system of concepts related in such a way that to understand any one of them you have to understand the whole structure it fits” (Fillmore 1982, 111).

L’Homme (this volume) provides an overview of the contributions of Lexical Semantics to Terminology, including knowledge-driven models and Frame Semantics.

3.6 Terminological variation

Grön and Bertels (2018) focused on term variation in the biomedical domain, more specifically on a corpus of Electronic Health Records. Distributional analysis showed that different sections of these clinical records varied systematically with respect to their lexical, terminological, and semantic composition, as well as their potential for term variation (Grön and Bertels 2018).

4. Challenges of distributional analysis and issues raised by specialized corpora

This chapter has revealed that distributional analysis in specialized corpora has a number of issues, mainly with respect to (limited) corpus size, multi-word terms, and specialized vocabulary. The following sections describe these challenges and discuss possible solutions.

4.1 Corpus size and data sparseness

A major problem for distributional analysis in specialized corpora is the limited corpus size in specialized domains, which can range from a few hundred thousand tokens to a few million. For general language, vast corpora can easily be built by using web crawling. For specialized domains, however, the availability of domain-specific texts is more problematic and inherent to the domain. As pointed out in Section 3.1, although large biomedical corpora of PubMed publications can be compiled in English, there are fewer texts available in other specialized domains. This is evidently a challenge for English, but it is an even bigger challenge for other languages.

Not only is corpus size an issue in specialized domains, but also data sparseness. Distributional analysis relies on vector space models, which suffer from data sparseness problems, and the high dimensionality of the co-occurrence matrix. In the case of smaller (specialized) corpora, the data sparseness problem becomes worse because of the limited number of occurrences and shared contexts or co-occurrences.

In order to overcome these problems, Périnet and Hamon (2015) developed an approach based on the abstraction of distributional contexts as a way to reduce the dimensionality of the co-occurrence matrix and thus overcome the data sparseness problem. They considered both single-word and multi-word terms. In their approach, context abstraction was based on semantic relations (hypernymy and synonymy), acquired from a training corpus, and involved methods of lexical inclusion, syntactic patterns, and terminological variation. This context abstraction approach was tested on a small medical corpus in French (+/- 84,000 tokens) and on a small medical corpus in English (+/- 178,000 tokens). It produced better results, because semantic groupings based on context abstraction were semantically more homogeneous (Périnet and Hamon 2015). This approach may lead to a more efficient use of distributional analysis in specialized corpora.

4.2 Multi-word terms and compositionality

Domain-specific terminology is characterized by a vast majority ($\pm 85\%$) of compound terms (Nakagawa and Mori 2003). Above all, the acquisition of synonyms of multi-word terms from specialized corpora is challenging. Previous studies have relied on external resources and dictionaries for the task of synonym extraction. Nevertheless, for the synonym extraction of multi-word terms, the use of dictionaries can often be deceiving.

To address this problem, Hazem and Daille (2015) developed an unsupervised semi-compositional method based on the combination of a distributional approach (i.e., distributional analysis) and a compositional approach (i.e., the compositional properties of multi-word terms). Semantic compositionality allowed for the acquisition of synonyms of multi-word terms from a specialized corpus (e.g., *énergie eolienne* and *courant éolien*). The unsupervised semi-compositional approach yielded satisfactory results for a specialized comparable corpus of renewable energy texts ($\pm 300,000$ tokens) in French, English, and Spanish. Results were also promising for a specialized comparable corpus of breast cancer texts ($\pm 200,000$ tokens) in French and English (Hazem and Daille 2015).

A recent study of a specialized corpus of 5 million tokens in the biomedical domain of stem cells, in English, focused on domain-specific concepts, expressed as multi-word terms (Brunet, Ferret and Tanguy 2020). Distributional count-based models were built from syntactic dependencies. The main challenge was to align the multi-word terms with the output of the dependency parsers.

4.3 Specialized vocabulary and in-domain knowledge

Specialized corpora consist of terms designating specialized concepts and expressing specialized knowledge. This is an important challenge for the use of Deep Learning models on such corpora because word embeddings are usually pretrained on large general language corpora. In the case of general corpora, off-the-shelf word embeddings yield satisfactory results and have become the main NLP approach in distributional analysis. However, most off-the-shelf word embeddings do not perform well on specialized corpora. Furthermore, training word embeddings or building word representations on small specialized corpora is problematic because of the limited availability of in-domain data.

In the clinical domain, El Boukkouri et al. (2019) used embedding strategies based only on general-domain resources. They combined off-the-shelf contextual embeddings (ELMo) with static Word2Vec embeddings trained on a small clinical corpus, which outperformed word embeddings learned from a larger medical corpus (El Boukkouri et al. 2019).

In 2019, a new pre-trained language model was made available to the scientific community. Bidirectional Encoder Representations from Transformers for Biomedical Text Mining (BioBERT) is a domain-specific language representation model which has been pre-trained on large English biomedical corpora (Lee et al. 2019). BioBERT is a BERT-based model for deep learning for use on English corpora in the (bio)medical domain, for various applications, such as Named Entity Recognition (NER). A similar contextualized language model, called CamemBERT, has been trained and optimized for French (Martin et al. 2020). An ensemble method consisting of CamemBERT and CamemBERT-bio (trained on French biomedical texts) yielded satisfactory results on an NER task on French biomedical texts in “Défi Fouille de Textes”, an Information Retrieval challenge (Copara et al. 2020).

5. Conclusions

This chapter started with a detailed explanation of the theoretical and methodological principles of Distributional Semantics which rely on the idea that words or terms with similar distributions have similar meanings. Different distributional semantic models (DSMs) follow the same four-step procedure, i.e., extraction of co-occurrences, weighting, dimension reduction and calculation of similarity, but with varying parameters. Collocates can be window-, syntax- or document-based and can be extracted at type- or at token-level. In terms of how the context information is stored, DSMs are either count-based or prediction-based. Count-based models are referred to as vector-based models, whereas prediction-based DSMs (e.g., Word2Vec) rely on neural network algorithms and give rise to neural embeddings or word embeddings.

Furthermore, this chapter provided an overview of research on distributional (semantic) analysis in specialized corpora, with interesting results in several specialized domains and for various terminological applications. Distributional analysis has been very productive for terminology extraction and domain-specific vocabulary expansion in the biomedical domain, especially for English. It has also been used for the (semi-)automatic creation of distributional thesauri, the extraction of semantic relations and knowledge patterns, and the identification of terminological variation.

Finally, this chapter discussed a number of issues raised by the use of specialized corpora, mainly with respect to corpus size, multi-word expressions and specialized vocabulary. In order to overcome the data sparseness problem, an approach based on the abstraction of distributional contexts has been used for dimensionality reduction of the co-occurrence matrix. To address the challenge of multi-word

expressions, the combination of a distributional approach and a compositional approach has yielded satisfactory results. The issue of specialized corpora has been addressed by pre-training domain-specific language representation models on large specialized corpora (e.g., BioBERT and CamemBERT).

Some distributional analysis problems have been solved or are well on their way to a satisfactory solution. Especially, in the biomedical domain, there have been major advances. However, some issues are still pending and may offer challenging future research possibilities and opportunities for distributional analysis in specialized corpora, or even in general language corpora.

Thanks to the use of distributional semantic models (DSMs), distributional analysis has adopted a bottom-up approach. In other words, it relies on the distributional properties of words and/or terms observed in large text corpora and converts them to semantic properties. As such, DSMs provide a rather coarse-grained representation of lexical meaning (Lenci 2018). It is still not clear whether statistical co-occurrence information, even though it is based on very large data sets and on deep learning, is sufficient to reliably calculate semantic information. More specifically, it remains unclear what kind of semantic information is provided and which semantic information remains inaccessible. This issue had already been raised by various researchers (e.g., Sahlgren 2008; Lenci 2008; Fabre and Lenci 2015) and is still pending.

Another related issue is the question whether shared distributional properties display semantic similarity or semantic relatedness (Harispe et al. 2015; Lenci 2018). Two words sharing the same co-occurrences can be synonyms, antonyms, or co-hyponyms but they can also be thematically associated (e.g., *patient* and *doctor*). Semantic similarity only involves taxonomic relationships, although some studies based on DSMs and semantic measures also consider other relationships to be semantically similar. Generic lexical representations could be further enriched and fine-tuned in order to take language variability into account or specific language constraints (Popa 2019).

Future research will also have to deal with the issue of lexical ambiguity and polysemy in (specialized) corpora, for example by means of token-based DSMs (Heylen et al. 2015) to account for token-level usage contexts and to gain insight into the related senses of polysemous words and terms.

As deep learning techniques with neural network models and word embeddings become the mainstay in distributional analysis on general language corpora, they will probably find their way into specialized domains, as they already have in the biomedical domain, with a domain-specific word-embedding model. Nevertheless, the availability of (very) large specialized corpora and in-domain knowledge sources still remains a huge challenge.

PART 6

Terminology and Cognitive Linguistics

Units of understanding in Sociocognitive Terminology studies

Rita Temmerman

Vrije Universiteit Brussel

In Sociocognitive Terminology theory the point of departure is the ‘unit of understanding’ (UoU). A UoU exists because of three key elements. First, the understander relies on the senses to get a grip on aspects of reality. Second, the understander has a functioning mind. Third, the understander has proficiency in at least one human language allowing him to put into words his personal understanding of a unit as well as to understand discourse that represents earlier references to the UoU under consideration. Four more aspects need to be added to make this theory complete: prototypicality in understanding, embeddedness in society, a diachronic perspective, and the impact of metaphorical framing. Three case studies that were executed within the framework of Sociocognitive Terminology studies are presented.

Keywords: Sociocognitive Terminology studies, unit of understanding, embodied understanding, situated understanding, understanding discourse, cultural understanding, negotiated understanding, diachronic understanding, metaphorical understanding

Believe me, I speak as my understanding instructs me
–William Shakespeare. *Winter’s Tale* I:1

1. Terminology studies: From prescriptive to descriptive

We have adopted the term ‘unit of understanding’, which was originally introduced in Sociocognitive Terminology theory (Temmerman 2000) to deal with the inadequacy of classical concept theory for most specialised fields. The unit of understanding pertains to the understanding of a referent in a situated reality and how this understanding is expressed in language (terminology in the discourse of a subject-field). The ongoing development of units of understanding can be

explained as the result of several simultaneously active factors: (i) the urge specialists have for more and better understanding; (ii) the interaction between language users (specialists in different disciplines) to describe units of understanding related to a specific subject area (e.g., biology, informatics, law); (iii) prototype structure in the understanding of categories which can be seen simultaneously as the result of and as one of the causes of knowledge evolution; (iv) cognitive models playing an important role in the development of new ideas. Early endeavours to establish a General Theory of Terminology (GTT) or a science of terminology (Wüster 1979) need to be situated in the context of standardization (see Chapter 1 by Humbley in this volume). Standardization is not in the first place about language. A standard could be seen as a formula that describes the best way of doing something for people such as manufacturers, sellers, buyers, customers, trade associations, users or regulators. As standards are drawn up in a natural language (e.g., English, French, Chinese) and they contain terminology, ISO/TC 37, a technical committee of the International Standardization Organization, was put into operation in 1952 in order “to find out and formulate general principles of Terminology and Terminological Lexicography” (as terminography was called at that time). A standard on terminology saw the light with ISO 1087:2019 entitled *Terminology work and terminology science – Vocabulary*. In the 2019 version of ISO 1087, ‘concept’ is defined as “unit of knowledge created by a unique combination of characteristics” (ISO 2019, 3.2.7). Two notes are added that somehow undermine the definition of ‘concept’. The first note states that concepts are not necessarily bound to particular natural languages. “They are, however, influenced by the social or cultural background which often leads to different categorizations”. This note undermines “unique combination of characteristics” found in the definition. The second note is quite incomprehensible, even though it is meant to clarify the definition of ‘concept’: “This is the concept ‘concept’ as used and designated by the term *concept* in terminology work (3.5.1). It is a very different concept from that designated by other domains such as industrial automation or marketing”. Apparently, the concept of ‘concept’ is hard to define.

However, it soon became apparent that “a term is a name for a concept” was inappropriate, since the structure and function of language as a means of expression and communication was overlooked, as well as the creative force of language in thought processes and argumentations. Several scholars with a background in Applied Linguistics reacted to the one-sided approach of ‘terminology science’ as defined in ISO (2019). These experts were studying the terminology of specific domains in discourse, often from a multilingual perspective as most of them were active in the training of translators, interpreters, and multilingual communicators (see e.g., Johnson and Sager 1980; Sager 1990; Cabré 1999c; Gaudin 1993; Temmerman 2000; Kageura 2002; Humbley 2003; Collet 2004; Diki-Kidiri 2008;

Kageura and L'Homme 2008; Faber 2009, 2015; Picton 2011; Pecman 2012b, 2014; Peruzzo 2014; L'Homme 2018; Rossi 2018). They made attempts to give terminology studies a truly scientific grounding.

Research questions were formulated, and research methodologies were proposed inspired by e.g., Corpus Linguistics, Frame Semantics, Cognitive Semantics, Text Linguistics, Discourse Analysis, Special Language Studies, Cultural Studies, and Translation Studies. The theoretical assumptions of the GTT were questioned and reconsidered. The overemphasis on the ideal term for “a clearly defined concept” was corrected. Previously overlooked aspects like e.g., metaphorization, variation, terminology development over time, were given attention.

Sager (1990) and Cabré (1999c) emphasized the communicative function of terminology in “effective communication”. In 1980, Johnson and Sager had already emphasized the relative flexibility of reference. Relativity was also dealt with by Collet (2004) when she wrote her influential article entitled “What will be considered a term?” Collet chose to define the term within the theoretical framework of Text Linguistics. As she sees it, a term has two fundamental functions in a text for specific purposes: its naming function and its cohesion and coherence producing function. In doing so the emphasis is on the term’s original and dynamic contribution to text cohesion and to text coherence, two characteristics of a term that were not considered in the GTT. There are many more examples of how terminology studies were given a more scientific approach, several of which are treated in the other chapters of this volume.

2. The extended semantic triangle of understanding

Sociocognitive Terminology (Temmerman 2000) starts from a reflection on meaning, reference, thought, and understanding. The semantic triangle as dealt with by Richards and Ogden ([1923] 1989) is revisited. Their theory is a contextual theory of signs. They argue that the relation between a word (symbol/form) and a thing (referent in the world) it refers to is not direct. It is mediated by a concept (reference/thought in the mind). A semantic whole is expected to include the three parts in a context: symbol or word in language, thought or concept in the mind and referent or thing that can be observed in the world. Symbols are “those signs which men use to communicate one with another and as instruments of thought, occupy a peculiar place” (23).

For Richards and Ogden, discursive symbolization involves “weaving together of contexts into higher contexts” ([1923] 1989, 220). For a word to be understood “requires that it form a context with further experiences” (210). Figure 1 represents an extended understanding of the semantic (also referred to as semiotic) triangle.

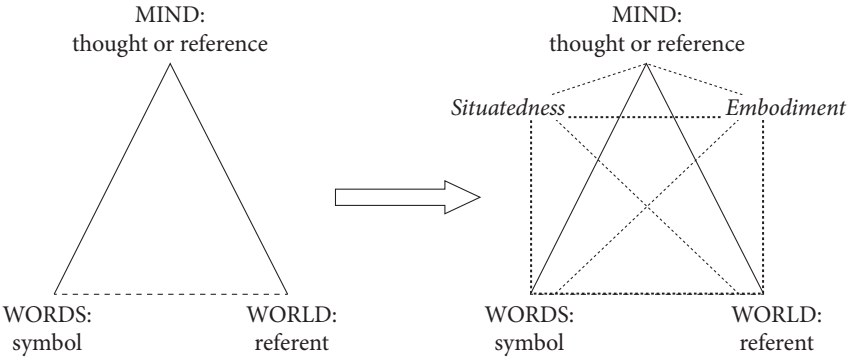


Figure 1. The semantic triangle (on the left) represents understanding. A thought or reference in the mind is linked to a symbol (words in natural language) and a referent in a reality (the world). Two elements are added in the extended semantic triangle (on the right): *Situatdness* (understanding occurs in a situated environment) and *embodiment* (people understand aspects of reality via sensory observation: Seeing, feeling, tasting, hearing, smelling)

Sociocognitive Terminology studies take this semiotic triangle as a starting point to further develop a dynamic model of understanding (Figure 2). There are several dimensions to understanding. First there is the linguistic dimension. As Firth (1957, 11) wrote, “You shall know a word [*term*] by the company it keeps”. The syntagmatic aspect of understanding a term can be studied in texts. A second dimension concerns understanding as categorization in the mind. This understanding can be the result of observation via sensory experience. Johnson (2015) sees human understanding as profoundly embodied. It is rooted in how our bodies and brains interact with, process, and understand our environments in a way that recruits bodily understanding. Thirdly understanding is the outcome of communication and negotiation in a situated environment (the *situatdness* of understanding). Putnam (1975) discusses the dynamics of understanding and refers to communication and negotiation as “the division of linguistic labor”.

Whenever a term is subject to the division of linguistic labor, the ‘average’ speaker who acquires it does not acquire anything that fixes its extension. In particular, his individual psychological state certainly does not fix its extension; it is only the sociolinguistic state of the collective linguistic body to which the speaker belongs that fixes the extension. (Putnam 1975, 229)

This view on understanding as socially distributed is also found in Lindblom (2015) (see Section 4.3.).

Understanding is a process. Knowledge is the result of a continuous dynamics of understanding, misunderstanding, and deeper understanding. We can study

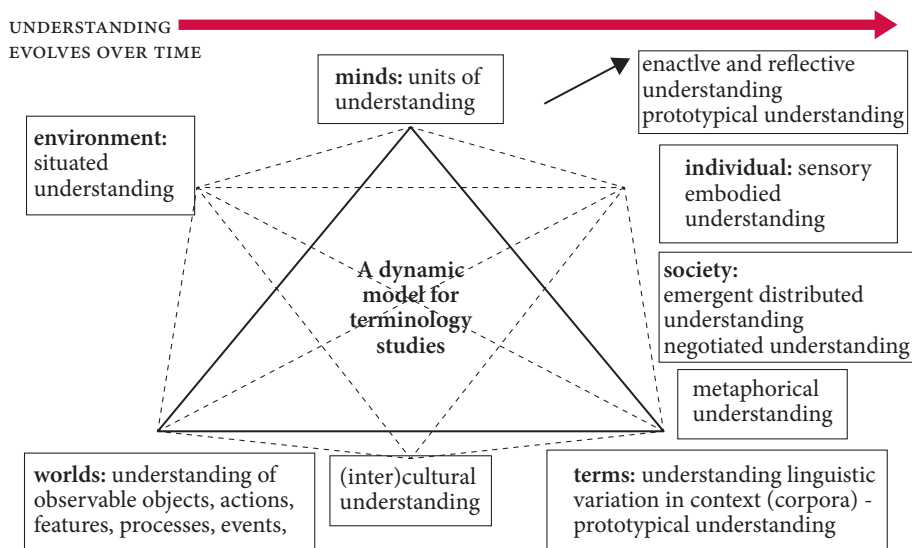


Figure 2. This dynamic model for terminology studies further develops the extended semantic triangle in Figure 1 (situated and embodied understanding). Several supplementary approaches for studying terminology and understanding are added: Understanding is prototypically structured (which plays at the level of terms in language and of the comprehension of new units); understanding is (inter)culturally determined; metaphorical understanding; understanding is emergent in society, it is distributed over several individuals who can negotiate understanding and ‘come to terms’ about understanding; understanding evolves over time

this process by observing the cognitive potential and the development of natural language(s) and more particularly of lexical items, i.e., terminology, in specialized domains. Renewed interest in both the dynamics of cognition and the creative potential of language has shifted the perspectives of terminology studies to the creation of neologisms in special languages, the study of variation, research concerning ambiguity, synonymy, metaphor, phraseology, the role of diversity, variation, polysemy, vagueness and indeterminacy, and tentativeness in human understanding.

3. A dynamic model for Sociocognitive Terminology studies

In order to illustrate the dynamic model for terminology studies (Figure 2) we first introduce an example of a case study on the dynamics of understanding and legal neologisms as discussed in Kerremans, Temmerman and Andries (2016). It describes the process towards legal neologisms in French and Dutch in Belgium, starting from the legal definition of ‘mother’ in Belgium. The observation is that – as

a consequence of scientific and societal developments – three new UoUs challenge the legal definition: surrogacy motherhood, anonymous motherhood, and co-motherhood.

According to Belgian law, the legal mother is the woman who has (i) given birth to the child and (ii) whose name features on the child's birth certificate. A surrogate mother is a woman who meets both legal conditions but lacks the intention to take up motherhood herself. Anonymous motherhood arises when a woman gives birth to a child while her name does not feature on the birth certificate, as yet an illegal form of motherhood in Belgium. A woman who intends to be the child's mother because of her intimate connection with a child's mother, but she has not given birth to the child, nor is she mentioned on the birth certificate, is known as a 'co-mother'. Following a process of parliamentary debates, as soon as the parliament approves (accepts by majority of votes), legal neologisms become legal terms in Dutch, French, and German, the official languages in Belgium.

For this case study a research corpus of all law proposals ever deposited concerning these three types of motherhood was compiled. The corpus consists of white papers and preparatory documents challenging the classical legal definition of 'mother'. The dynamicity of legal units of understanding was analyzed both from a category-specific and a denominative angle by studying the Belgian bills and white papers, written by policy makers in collaboration with legal specialists of the Dutch and French language communities. Evidence was found that language-specific understanding may persist throughout categorization within the same legal space and that denominative variation may be linguistically and culturally inspired. The contrastive study (French – Dutch) was based on close reading of all proposals submitted by French-speaking and Dutch-speaking members of the federal parliament on the new understanding of motherhood.

By executing first, a UoU approach (what is the core definition of each type of motherhood?) and then a denominative approach, we found an answer to the question of which legal neologisms are used to discuss the UoUs that both language communities have in common. The denomination of a UoU is a word, a word group, a phrase or sentence referring to a UoU. These are terms with 'co-referential status' related to the UoU in 'lexical chains': "cohesive ties, sharing the same referent, lexically rather than grammatically expressed" (Rogers 2007b). This method yields definitions and how they are questioned, neologisms and what they refer to, and variation related to a UoU in texts.

What is demonstrated with this example is that to describe units of understanding related to law making, Sociocognitive Terminology studies term behaviour in specialised texts. Collet (2004, 109) defines a term as "a semantically charged linear structure, which names an abstract or concrete reality within a special subject field". In Sociocognitive Terminology, terms are held to be the starting point in

terminological analysis. They can be studied in text corpora and beyond (see case studies in Section 4 and conclusion), and at least three types of contexts must be taken into consideration: the lexical context, the situational context, and the cognitive context. Sociocognitive Terminology shifts the perspective from symbols and meaning making to understanding. The shift from meaning to understanding has important implications for terminography. Terminography incorporates several practical tasks, such as the identification and analysis of terms in specialized texts as well as the organization and description of these units in terminological resources. Focusing on understanding first, emphasizes the fact that there are different degrees of understanding, depending on the type of user and other case-related factors.

The example of motherhood in Belgium illustrates the dynamic nature of units of understanding because understanding reality and certain phenomena in reality continuously undergoes changes. The web of understanding woven in context assimilates new contents or accommodates itself to them. Moreover, understanding occurs across a continuum: we can understand to differing depths. Perkins (1996) believes that from the point of view of the understander, understanding is the ability to think and act with what one knows. As he puts it, an understanding of a topic is a “flexible performance capability” (Perkins 1996, 40).

An expert understanding a topic of study will be able to process a unit of understanding by e.g., finding examples, generalizing, analogizing, representing in a different way, etc. This flexible performance capacity concerning the UoU shows one’s understanding and advances the understanding which is often expressed in language. Perkins (1996) calls these *performances of understanding*. His performance view of understanding contrasts with another view of understanding prominent in both our everyday language and in cognitive science. We often think of an understanding as a representation or image or mental model that people possess rather than of their “performance capabilities”. The performance view of understanding challenges the centrality of representations. What the domain specialist understands is not just a representation but a performance capability (see also Section 4.1).

As the case of motherhood in Belgian law-making illustrates, four more aspects need to be added to make the dynamic model for terminology studies complete. The first one is that a unit of understanding is prototypically structured (Geeraerts 1988). The second one is that understanding is embedded in society and occurs in a cultural environment. Members of a cultural group negotiate and renegotiate their emergent cultural cognition across time and space (Sharifian 2008). The third one concerns a diachronic perspective on how understanding evolves, and the fourth concerns the impact of metaphorical framing on understanding and term creation (neologisms) (Temmerman 2008). We will discuss each of these four aspects in what follows.

3.1 Prototypical understanding

Categorization in the mind is a dimension of understanding. Wittgenstein (1953, cited in Rosch 1978), argued that the referents (exemplars) of a word need not have common elements to be understood and used in the normal functioning of language. A family resemblance might be what links the various referents of a word. Each exemplar has at least one, and probably several, elements (features) in common with one or more items, but no, or few, elements are common to all items (Rosch and Mervis 1975, 574–575). This phenomenon, which was largely studied in psychology, is known as prototypical understanding of our environment and of language.

The prototype-based conception of categorization originated in the 1970s with Eleanor Rosch's psycholinguistic research into the internal structure of categories. The results of the research by Rosch (1978) initially relate to perceptual categories (colour terms). Rosch observed that the tendency to define categories in a rigid way, clashes with the actual psychological situation. Perceptually-based categories do not have sharply delimited borderlines. Instead of clear demarcations between equally important conceptual areas, one finds marginal areas between categories that are only unambiguously defined in their focal points. Rosch developed this observation into a more general prototypical view of natural language categories, more particularly, categories naming natural objects.

The theory implies that the range of application of such categories is concentrated round focal points represented by prototypical members of the category. The attributes of these focal members are the structurally most salient properties of the concept in question, and conversely, a particular member of the category occupies a focal position because it exhibits the most salient features. The most typical instance of the category 'bird', for instance, was the robin. Next came the sparrow, the blue jay, the bluebird, and the canary. At the lower end of the prototypicality scale chickens, penguins, and peacocks are found. Prototypicality implies that categories exhibit degrees of typicality; not every member is equally representative for a category. Most importantly, prototypical categories cannot be defined by means of a single set of necessary and sufficient attributes (Geeraerts 1988).

3.2 (Inter)cultural understanding and multilingual understanding

Sociocognitive Terminology studies also found inspiration in culture studies. Cultural Linguistics offers a theoretical and analytical framework that focuses on examining understanding that is rooted in the cultural experiences of speakers. The basic premise underlying the approach of Cultural Linguistics is that certain

features of human languages are entrenched in cultural conceptualisations such as cultural schemas, cultural categories, and cultural metaphors (Sharifian 2017). This approach offers a multidisciplinary understanding of cognition that moves beyond the level of the individual mind and its associated notions, such as ‘mental representations’.

Cultural cognition is a form of enactive cognition that is formed because of interactions between individuals across time and space. Crucially, cultural cognition is not equally shared by speakers across a speech community but is a form of distributed cognition. Speakers show variations and differences in their access to and internalisation of cultural cognition. Moreover, cultural cognition is dynamic in that it is constantly being negotiated and renegotiated across generations and through contact with other speech communities. This understanding of cultural cognition is entirely different from the essentialised notion of ‘culture’ that is often associated with linguistic relativity in the Whorfian tradition (Whorf 1956).

Whereas the traditional approach to Terminology Theory stipulates that culture should be excluded when building terminology and that one term in a language can only have one denomination in another language a new approach named Cultural Terminology developed in Africa. It originated in the problem encountered when producing school materials for African children. How can Africans acquire knowledge when being educated in their own languages without taking their culture into consideration? Marcel Diki-Kidiri, a linguist specialized in the study of the Sängö language, is a leading figure in the studies of the development of African languages. He initiated a Cultural Terminology approach to language development (Diki-Kidiri 2008, Chapter 9 in this volume) and published on how to introduce and durably sustain less resourced languages in cyberspace, how to revitalize and modernize them, and how to develop a language teaching policy in a multilingual context. Cultural Terminology in Diki-Kidiri’s approach does not seek to harmonize the vocabularies between languages but rather to make the languages progress by involving the speakers of the language in finding the terms.

3.3 Understanding evolves across a continuum

The process of understanding involves tentativeness (trying to put to words) and degree. By degree we mean that the subject at first understands more or less, later in a more profound way or differently, with growing insight. The understander may be capable at one point of rejecting a previous understanding and replacing it with a new one.

Since understanding is embedded in society, sociocognitive processes in humans describe the ways individuals draw inferences about other people’s units of

understanding and terminology and how they weigh social situational factors in making these inferences. Members of a cultural group negotiate and renegotiate their emergent cultural cognition across time and space (Sharifian 2008). From a careful study of a body of texts which were published over a period of time, an evolution can be observed from imprecise to more precise designation (see e.g., Temmerman 2008).

3.4 Metaphorical framing and understanding

Through metaphor, people consciously and unconsciously create images, represent and symbolise ideas, and articulate understanding of phenomena or situations (Lakoff and Johnson 1980). The role of metaphor in human communication and cognition is relevant in terminology studies. Metaphor is not just a linguistic phenomenon. In metaphor creation by specialists, various cognitive processes, such as perception and imagination and language skills interact.

Based on the study of scientific terminology in molecular biology and in invasion ecology, the role metaphorical framing plays in new understanding was illustrated (Temmerman 2014, forthcoming). Whereas conceptual metaphors of creative thinking generate novel insights and neologisms, misappropriation may occur with ideologically inspired emotional metaphors.

Accurate scientific data paired to conceptual metaphors appear to be the 'first frame' for scientific debate whereas intentional emotional metaphors can be a powerful force in shaping a 'second frame' of personal ideology and self-interest in an attempt at influencing policy decisions. Over the last two decades the observation and description of metaphorically framed terminology in corpora of specialized discourse has gained ground (e.g., Ureña and Faber 2010; Demaecker 2017; Rossi 2017, 2018).

Having explained the tenets of the dynamic model for terminology studies as visualized in Figure 2 we will sketch how this model was used to describe terminology in different disciplines in Section 4.

4. Case studies

Within the framework of a theory of understanding, several aspects of terminology can be studied and described. Hypotheses concerning the dynamics of terminology can be confirmed or refuted based on observations in a representative corpus of discourse and by surveying and specialist consultation. We present three case studies that were executed within the framework of Sociocognitive Terminology studies.

The first case bears on a UoU characterized by vagueness in texts produced by the European Union bodies (4.1). The second case deals with UoUs in molecular biology (4.2). In a third case study (4.3) UoUs of sensory experience in food sciences are used to illustrate the impact of embodied understanding on terminology creation within the domain specific languages of food and wine tasting and describing.

4.1 Vague units of understanding in the European Union

This study zooms in on the emergence of European cultural awareness and its effect on the language habits of the European Union. Within the European Union (EU), collective cognition is constructed and widely discussed leading to new units of understanding and terminological neologisms in Eurolanguage (Biel, Biernacka and Jopek-Bosiacka 2018). These units of understanding are part of the European emerging common frame of reference towards a European legal culture (Temmerman 2018).

In general, EU documents are meant to represent the collective voice of all individual Member States. Vagueness allows these Member States to agree on measures such as the establishment of a fund for ‘integration’ without sacrificing their respective issues, goals, and priorities.

A study on the European understanding of the UoU ‘integration’ is conducted by a triangulation of results obtained by (i) surveying young EU citizens and third-country nationals living in Brussels; (ii) consulting publications of legal specialists; (iii) analyzing the discourse in official European documents and (iv) close reading and performing a terminological analysis of knowledge rich text fragments (Meyer 2001 see Chapter 5 in this volume).

4.1.1 A floating signifier: ‘Integration’

As a word, ‘integration’ is a so-called internationalism that finds its origin in Latin. It is polysemous in the English language, as are its equivalents in other languages. It is a so-called floating signifier, i.e., a symbol loose enough to mean many things to many people, yet specific enough to invigorate action in a particular direction. Some examples of the polysemy of ‘integration’ will be given. First, in engineering, ‘system integration’ is the process of linking together different computing systems and software applications, physically or functionally, to act as a coordinated whole (Wikipedia 2020).¹ Second, ‘integration’ in education requires adjusting to an education system; students with disabilities are integrated, which means educated partially in a special education program (UNESCO 2017). Third, ‘integration’ of

1. Wikipedia entry for ‘system integration’ (https://en.wikipedia.org/wiki/System_integration) Accessed 11 December 2021.

immigrants is defined by the Unesco Glossary (2017)² as a synonym of acculturation, i.e., the process of making someone equal or fitting to the rest of society.

In what follows, several elements of the sociocognitive model of understanding are used to analyze the vagueness of this third meaning of the word ‘integration’ as it becomes a UoU in the institutional context of the EU and of one of its member states, Belgium.

4.1.2 *A situational context for the UoU ‘integration’ in Brussels (2017–2019)*

Understanding happens within a personal situational context. In a written survey held among exchange students in Brussels in 2018. Hannah, a USA citizen who had enrolled in a master’s program at Vrije Universiteit Brussel, testifies about her personal process of understanding, which we might see as her experienced performance in understanding (Perkins 1996) ‘integration’ in Belgium:

I received my first Belgian visa in April 2017. When I opened my passport, a paper fell out which had been placed there by the Belgian consulate:

Willingness to integrate in Belgian society – New general condition of stay from 25.01.2017

The efforts towards integration in Belgian society by a foreigner who is admitted to or permitted to a stay of more than 3 months in Belgium will henceforth be verified by the Immigration Office (FPS Interior). We strongly recommend that you consult the website of the Immigration Office for a detailed explanation regarding this new general condition of stay.

At that time, Hannah did not understand what was meant by “efforts towards integration in Belgian society”. As she was strongly recommended to consult the website of the Immigration Office, she looked for more information there, but the website did not provide more details at that time. In her personal testimony she writes:

I kept that slip of paper and wondered how I would prove I had made an effort to integrate in Belgium – a country I had never been to, whose languages I did not know, and whose citizens I had never interacted with. I filed the note with other important documents, but after a few weeks of navigating bureaucracy in Belgium I stopped worrying about a potential visit from immigration officers, who never did contact me. I am now on my second Belgian visa. I found the same notice tucked inside my passport – I showed my father in California and had a good laugh before tossing it in the trash. My actual words were, “I speak French, I like frites with mayonnaise, my favorite beer is Chimay and I ride a bike in Brussels. I’m pretty sure I’m integrated.”

2. UNESCO, Migration and Include Societies (<http://www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/integration/#topPage>) Accessed 11 December 2021.

What we witness here is Hannah's ability to think and act with what she knows already in order to try and understand the UoU 'integration' in a situational context. She furnishes proof of flexible performance capability (see Section 3) in her attempt at understanding what integration means for her in Brussels in her situation.

4.1.3 *A process of understanding within the EU: 'Migrant integration' and 'third country national integration'*

According to the specialists in EU law-making, Garcés-Mascareñas and Penninx (2016) 'integration' was defined in a rather limited way until 2003 in the European Union. EU policies started from the implicit assumption that if the legal position of immigrants was equal to that of national citizens and if adequate instruments were in place to combat discrimination, integration processes could be left to societal forces. Thus, legal integration (= equality) was to be ensured by means of the directives on family reunification and free movement after 5 years and by comprehensive anti-discrimination directives.

In hindsight, the process of understanding 'integration' went through several phases: in 2003 integration was understood as a two-way process, 'based on reciprocity of rights and obligations of third-country nationals and host societies [and foreseeing] the immigrant's full participation' (Garcés-Mascareñas and Penninx 2016, 1). In 2004, common basic principles (CBPs) were formulated by the Council of Ministers responsible for integration. The first article of the CBPs defines integration as "a dynamic, two-way process of mutual accommodation by all immigrants and residents of Member States". Following that, National Contact Points on integration were designated by the ministries responsible for immigrant integration policy to promote information exchange, to monitor progress, and to disseminate "best practices" at the national and EU levels.

With a similar purpose, the European Integration Forum was established. This EU platform of representatives of civil society and migrant organizations was a forum for consultation, exchange of expertise, and identification of policy recommendations. Even more significantly, the European Integration Fund (EIF) was put in place to finance national programmes and community actions with a total budget of € 825 million from 2007 to 2013. In this process of growing insight, the EU has gradually expanded its definition of immigrants' integration.

A major shift in policy framing came in 2011 with the renewed European Agenda for the Integration of Third-Country Nationals, which added the countries of origin as a third key actor in the process of immigrants' integration, thereby introducing the three-way process (Garcés-Mascareñas and Penninx 2016).

4.1.4 *Regulation 516/2014 establishing the Asylum, Migration and Integration Fund*³

Regulation 516/2014 comprises the same term ‘integration’. It does not state explicitly what ‘integration’ means in the context of the European Union in 2014. Regulation 516/2014 is, essentially, an announcement that the EU will provide funds for Member States to integrate incoming third-country nationals (TCNs).

Most EU documents are meant to represent the collective voice of all individual Member States. Vagueness allows these Member States to agree on measures such as the establishment of a fund for ‘integration’ without sacrificing their respective issues, goals, and priorities. Regulation 516/2014 does not explicitly define integration anywhere in the twenty-seven-page document. The document uses language that closely resembles the definition of integration provided by the European Council: “a dynamic, two-way process of mutual interaction.” It does not specify which parties partake in this ‘two-way process’, though it can be inferred that this refers to Member State governments, their citizens, and the EU on the one hand, and people immigrating to EU Member States on the other (“third country nationals” EU-terminology). Regulation 516/2014 is, essentially, an announcement that the EU will provide funds for Member States to integrate incoming third-country nationals (TCNs).

By looking specifically at what areas the EU intended to fund, it is possible to infer what actions, qualities, or achievements the EU believes define integration without having to look beyond the source text. We can infer that ‘integration’ is defined as a process whereby people who have moved to a Member State of the EU from a non-Member State do the following: form connections with citizens, secure employment and housing, participate in and contribute to local culture, participate in democratic processes, benefit from the same access to public and private services as citizens of their host country, such as education and healthcare. The document also specifies that – ideally – integration begins prior to departure from one’s home country “through programmes such as job training, civic orientation courses that impart knowledge of the destination country’s culture and customs, and language courses.” Moreover, integration requires the participation of state, regional, and local governments, who should help with “housing, means

3. REGULATION (EU) No 516/2014 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL (<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0516&rid=1>). Accessed 11 December 2021. On 16 April 2014, the regulation establishing the Asylum, Migration and Integration Fund, amending Council Decision 2008/381/EC and repealing Decisions No 573/2007/EC and No 575/2007/EC of the European Parliament and of the Council and Council Decision 2007/435/EC.

of subsistence, administrative and legal guidance, health, psychological and social care, child care, and family reunification, as these support people in the process of immigration”.

All this information contributing to a tentative definition of ‘integration’ furnished proof that as a unit of understanding (UoU), ‘integration’ is both ambiguous and vague; ambiguous because its meaning is open to multiple interpretations, and vague in that because of its nonexistent definition in the document, the signifier lacks precision and detail. This is not the result of a linguistic oversight by the writers of Regulation 516/2014. The ambiguity and vagueness attached to integration are, in fact, intentional and even useful for the successful implementation of integration programmes in EU member states. Vagueness and indeterminacy serve functional interpretation. Countries as geographically, culturally, and economically distinct as Denmark, Spain and Poland will have different – even divergent – definitions of successful ‘integration’ in their respective social, economic, and political contexts.

If Regulation 516/2014 tied funding for Member States to the implementation of a specific programme structure to ensure the unilateral realization of an EU definition of integration (perhaps one that included the same measurable goals for employment, housing, civic engagement, school enrollment, language acquisition, etc. for all Member States), this would predictably fail because it would not take each country’s context into consideration. Regulation 516/2014 highlights the important role of context (Temmerman 2016) in giving concrete meaning to ‘integration’ through its prioritization of local organizations/authorities over state/EU actors in designing and implementing integration programs. It states that the role of the Asylum, Migration, and Integration Fund is to support “strategies specifically designed to promote the integration of third-country nationals at national, local and/or regional level,” which “should be implemented mainly by local or regional authorities and non-state actors, while not excluding national authorities” because local actors can “choose the measures most appropriate to their particular situation from a range of measures available.”

Vagueness as a phenomenon is “essential to the political construction of the EU” (Van Obberghen, Temmerman and Kerremans 2018, 77). Most EU documents are meant to represent the collective voice of all individual Member States. Vagueness allows these Member States to agree on measures such as the establishment of a fund for ‘integration’ without sacrificing their respective issues, goals, and priorities. It is also important to consider the power of cultural context in relation to Regulation 516/2014. The regulation is careful not to suggest that ‘integration’ means following a pan-European culture; rather, it speaks of supporting integration on a regional and local level into ‘receiving societies’, which is itself a vague unit of understanding.

4.1.5 *Integration in Belgium*

European Member States are left to decide for themselves how broadly or narrowly to define ‘society’, and at what level of their respective states to assign responsibility for creating EU-funded integration programs. In Belgium, for example, the term could be applied to the country as a whole, to Wallonia, to Flanders, or to the Brussels region. Each of these subgroupings have their own respective cultures that coexist within the larger ‘grouping’ of Belgium, which begs the question of how the EU’s call for integration of immigrants into their society is interpreted and implemented one cultural context versus the other.

Mandin (2014) discusses the divergent approaches to integration in Flanders and Wallonia. He points out that Flanders has a more cohesive, if not aggressive, “inburgering” programme which “promotes higher cultural conformity through an integration programme including Dutch courses, social orientation courses, and information about the values and norms of the Flemish society”, whereas Wallonia has the Dispositif d’Accueil des Primo Arrivant (“First-Time Visitor Reception”) programme, which is less comprehensive.

For Mandin (2014, 8) integration happens across the intersections of not just nationality, but also race and class. He writes, “in Brussels, the integration of the many foreign workers related to European institutions (often called “expat” for expatriate), who benefit from high education levels and better socio-economic resources, is rarely considered”. Expats may be third country nationals too e.g., foreign workers with good socio-economic resources from USA, China, Japan, etc.

4.1.6 *Purposeful vagueness of floating signifiers*

Floating signifiers, especially those that hold as much sociopolitical weight as ‘integration’, present a challenge to translators, interpreters, and multilingual communicators because of their vagueness and indeterminacy. These characteristics are at the heart of language. Making the effort to maintain the vagueness and ambiguity of polysemous terms not only preserves the integrity of the text or speech in question, but it also strengthens a translator or interpreter’s linguistic skills. Vagueness, ambiguity, and floating signifiers appear in all genres of text and speech. However, for those translators and interpreters working in the political and legal arenas, it is important to study the terminology carefully. A translator working for the European Commission could create confusion, at best, and a potential crisis at worst by adding over-specification to the UoU integration. This might have far-reaching ramifications.

4.2 Intercultural understanding in molecular biology in English and French

The second example of a study within the framework of the dynamics of Sociocognitive Terminology studies deals with the domain-specific language of molecular biology. In previous work (Temmerman 2014), a diachronic bilingual analysis of terminology development was executed, based on discourse analysis of selected publications in English and French. The study illustrated embeddedness of understanding in a cultural environment, prototypical understanding, and multilingual metaphorical understanding.

Precision in naming is often a problem because, initially at least, the UoU itself is not clearly delineated and consequently, during the process of discovering something new, a name is transitory. Few UoUs become delineated once and for all, since phenomena both natural and those based on human activities are in constant transition, and the ways of referring to them also undergo change.

Term formation occurs in a particular environment which influences the naming activity, e.g., a laboratory. At some point, when a new insight occurs, when a technique is developed, when a phenomenon is discovered, a new term may be coined. The subject specialists may create a term based on the intuitive meaning extension of an existing lexeme.

Temmerman (2014) deals with a series of developments in the history of molecular biology as a sequence in time. Several UoUs regarding the presumed reality of protein synthesis in eukaryotic cells were described as they developed within the reasoning framework of molecular biologists. We demonstrated how the creative power of language functions in neology creation in discourse, and how terminological variation facilitates, enhances, and specifies the understanding of a presumed reality.

Molecular biologists exchange ideas with peers and other players through conversations, oral presentations, and scientific publications, thus participating in the scientific debate. Peers may prove a researcher wrong and point out misunderstandings or alternative understandings of terminology, followed by reconsiderations, further reasoning, further experiments, and tests in order to reach a more profound understanding. Most of the time, peer response consists in the careful reformulation of ideas put forward by their colleagues. Nuances will be expressed in a tentative manner. To coin terms and then to reformulate their meaning in variation across grammatical categories is a way of getting a better hold of new units of understanding; it is the way in which we transform our experience into knowledge (Halliday 1999, 2004). For Pecman (2014, 21, see also Chapter 12 in this volume by Pecman and Kübler), “neology ‘happens’ most frequently in discourse, and it is very closely linked with the construct of argumentation or a new knowledge unit”. This knowledge unit is what we refer to as a unit of understanding (UoU).

The increasingly advanced understanding of protein synthesis and mRNA splicing is a dynamic process, leading up to “our current understanding of the dynamics, composition and structure of the spliceosome” (Anthony and Pomeranz Krummel 2013, 1). By conducting discourse analysis of publications written by domain specialists, we observed how UoUs are expressed in both English and French, and how both languages develop their own terminology.

French had relatively little interference from English in the domain of molecular biology until a few decades ago. DNA (deoxyribonucleic acid) was and is consistently named ADN (*acide désoxyribonucléique*) in French, RNA (ribonucleic acid) is ARN (*acide ribonucléique*) and mRNA splicing is called *épissage mARN*. The French-speaking world has a tradition of trying to find French words for all phenomena.

Until recently, French molecular biologists had not found it too difficult to create French neologisms as part of the process of the advancement of knowledge without having to resort to the influence of foreign language such as loan words. However, this has changed in recent times with the discovery of e.g., snRNPs (1980), also called *snurps*, for ease of pronunciation (FR: *snRNPs* and *snurps*), and later e.g., with the spliceosome (1993) (FR: *splicéosome*), the snurportines (FR: *snurportines*), snoRNA (small nucleolar RNA), called in French *petits ARN nucléolaires*, (but the abbreviation *snoRNA* is kept in French), ORFs (open reading frames, also *ORF* in French), and UTR (untranslated region, also *UTR* in French).

In recent years, French-speaking molecular biologists borrowed and incorporated more and more English neologisms in what Sapir (1961) refers to as their *language habits*. English, the lingua franca of science, influences their language habits. Neologisms in molecular biology are first created and therefore motivated within the English language habits. We found several examples of English neologisms for which it appears to be difficult to create a French equivalent.

Temmerman (2014) focused on the process of understanding *mRNA splicing* and the discovery of the *spliceosome*. Her study also illustrated the process of neology creation in French and demonstrated the language and culture-boundness of terminology creation based on domain-metaphorical understanding, as well as the problems involved in finding French equivalents for abbreviated forms that were initially created in English. The longstanding tradition of ‘*aménagement linguistique*’ (language planning) in the French-speaking world is not to be overlooked in studies of neology creation.

4.3 Embodied understanding in food sciences

UoUs of sensory experience in food sciences are used to illustrate the impact of embodied understanding on terminology creation of the domain-specific languages of food and wine tasting and describing. In this case prototypicality in understanding, as well as the emergent social and cultural understanding in different languages and cultures can be observed and described along with metaphorical understanding.

Strikingly, more and more scholars in various fields of study are investigating in what ways humans depend on language to clarify and categorize their sensory understanding of food and how they negotiate and therefore build terminology by sharing and communicating their individual sensations. Descriptors of sensory experience play a role when attempting to objectify the world of tasting and neologisms are coined to name new products.

4.3.1 *Sensory experience from a Sociocognitive Terminology perspective*

In Temmerman (2017), we discussed several aspects of sensory terminology in food and beverage description and naming. Experiencing food and drinks is personal, culture-bound, and language-specific. We distinguished between the sensory side of experiencing foods and beverages (tasting, smelling, feeling, visual, and auditory observation) i.e., how words are chosen to discuss sensory understanding and vice versa, how descriptors for quality assessment of food and beverage are understood and how their meaning is constantly negotiated. We illustrated empirical heuristics, searching for linguistic clues in trying to understand the mechanisms of naming the understanding of food. We dealt with the marketing potential of a characteristic in white wine ('minerality') referring to olfactory, gustative, and tactile sensations and of a new international product name (*smoothie*). It appeared to be difficult to objectify the understanding of particular descriptors. Sensory descriptors and neologisms show significant cultural differences expressed in different languages.

The terminology of food and beverage tasting has a cultural twist in the sense that it is transmitted across generations within socially defined groups where it becomes part of the group members' identity and culture (regional dishes and wines, typical beers for a geographical area, coffee grown or roasted in different countries, etc.). The UoUs emerge and terms are coined first in the language(s) of the region or country where the food is typical. Whenever a particular product is sold outside the region where it originated, there may be a need for terminology in several languages spoken by groups of consumers having a different cultural environment and background.

Sensory analysis is carried out in the food industry to evaluate products and is performed by tasting panels. Professional tasters are expected to understand and assess the sensorial characteristics of products and to ensure their quality. They

need to work within a common frame of reference. Such a frame of reference relies on a classification system (an ontology) and a lexicon consisting of descriptors and definitions (Charette, Hooker and Stanton 2015). Overall, the results of the tasting process are to be not only as objective as possible; they are also expected to be as precise as possible. Language allows for degrees of precision. In language a taster of, for instance, a dish or a wine can indicate degrees of positive expressions that are more precise or in some cases more poetic (and full of connotations) than for instance the word *good*. Non-professional tasters expressing themselves in English, have a long list of positive descriptors at their disposal like: *amazing, appealing, appetizing, delicious, delightful, divine, enjoyable, enticing, excellent, exquisite, extraordinary, fantastic, heavenly, luscious, marvelous, mouthwatering, palatable, pleasant, pleasing, satisfying, superb, tantalizing, tasty, terrific, wonderful, yummy*, and many more.

As consumers we were taught to like or dislike in a linguistic and cultural environment and therefore, depending on where we were born and brought up, our list of descriptors may vary. Linguists and translation researchers ever since Whorf (1956) and Nida (1964) have pointed out that language shapes the understanding of our environment. Whorf believed that the language one speaks determines one's thoughts (Whorf 1956). We hypothesize that language not only shapes our thoughts but also shapes our tasting experience during the process of language acquisition in our sociocultural environment.

As human cognition is embodied, we are interacting with the world through the senses. But, as soon as we want to reflect and communicate about our findings (Krishna and Schwarz 2014), our sensory experience needs to be captured in linguistic expressions. So, we might say that expectation patterns when tasting food and drinks are culturally and linguistically determined because of previous experiences (Shapiro 2014) as we were taught to use words to express what we taste in a cultural environment.

Because language users negotiate the meaning of the words that they employ to describe their sensory experiences, they are all contributing to emergent and ongoing socially distributed understanding of descriptors for food and beverage tasting. According to the distributed cognition approach, cognition is a relational process in which meaning, and intentions are emergent products of social interaction. In most situations, they can be viewed as a kind of distributed phenomenon rather than as an individual private mental act (Lindblom 2015). Interaction in language use is essential for the understanding and creation of terminology in the food sector. This is reminiscent of Putnam's (1975) hypothesis of "division of linguistic labor" that highlights the specialized distribution of knowledge and meaning of sensory descriptors and food terminology and categorization.

Three aspects of his theory lead to a better understanding of the process of understanding and neology creation. The first one is that the members of society

collaborate to try and agree on the meaning of the words they use. The second that regardless of the first aspect, most language users will connect most words with vague and not quite identical 'stereotypes', which are however sufficient for ensuring mutual understanding in a great many instances. The third is that whenever doubt arises, the language community leaves the final judgment to those members of society that have been given status as 'experts' (Smith et al. 2013).

In line with Putnam (1975), Shapiro (2014), and Lindblom (2015), we consider new words (neologisms) for new products and descriptors for quality assessment as emergent creations of social interaction, a kind of distributed understanding rather than an individual private mental act. Depending on the context one will need a balance between the quest for objectivity and the quest for new experience. Objectivity is needed in a context where when one speaker uses a word or term, it should mean precisely the same for the interlocutor (in quality analysis and control). Yet, intersubjectivity may be relevant whenever we strive for innovation, using fantasy, out-of-the-box thinking, and linguistic creativity in our search for new experiences (e.g., in marketing and international trade).

In the case studies concerning the use of sensory descriptors for the tasting of food and drinks we observe two forces: a process of standardization (top-down) on the one hand and a process of harmonization (bottom-up) in which the division of linguistic labour plays on the other hand.

5. Conclusion and future challenges

In this contribution we summarized the tenets of Sociocognitive Terminology studies and visualized the dynamic model for terminology studies. We illustrated in several case studies that Sociocognitive Terminology takes the perspective of the understander. The unit of understanding (UoU) is the core element in the study of terminology in contexts. A UoU is dynamic, ever evolving, prototypically structured, embedded in a sociological reality, and under constant negotiation. UoUs appear in language in a variety of expressions: as terms, as descriptions, as definitions, and as tentative discussions. Terminology can be observed in discourse in a panoply of situational environments.

Methodologies for Sociocognitive Terminology research are adapted to the specific research questions and to the domain under investigation (e.g., aspects of law making, of scientific discoveries, of sensorial experiences). Interdisciplinarity is favoured, as triangulation may give rise to interesting insights. Many terminology projects are restricted to corpus analysis using computer tools. As a text corpus is not always sufficient as a context to understand realities, Sociocognitive Terminology combines corpus methods with methods that explore other modes

of cognition, such as observation and referential description or psychological experimentation combined with diachronic analyses.

In future research within the framework of a theory of understanding, more aspects of terminology can be studied and described. Hypotheses concerning the dynamics of terminology can be confirmed or refuted based on observations in a representative corpus of discourse and by surveying and specialist consultation.

Floating signifiers present a challenge to translators, interpreters, and multilingual communicators because of their vagueness and indeterminacy. Since these characteristics are at the heart of language, making the effort to maintain the vagueness and ambiguity of floating signifiers not only preserves the integrity of a text or speech, but it also strengthens a translator or interpreter's linguistic skills. Of course, vagueness, ambiguity, and floating signifiers appear in all genres of text and speech. However, for political and legal discourse, it is important to study vague and potentially ambiguous terminology carefully. More work needs to be done on the role of these phenomena in comprehension and communication.

There may be opportunities in Digital Humanities, a young field of research, focusing on the use of computational techniques and data management to support research in terminology and understanding. One of the key drivers of the field is the recent availability of large digital repositories within many areas in society. These repositories offer the prospect of applying computational visualization, querying and analysis techniques, leading to the discovery of new knowledge and insights. The field also applies new digital techniques to enhance approaches, processes, skills, and methods developed and applied in the humanities.

Digitized archives are not only useful for the preservation of data referring to the past but can also be of interest for studying understanding in the past. Analysing large archives, such as of newspapers and journals, can furnish answers to questions concerning the link between language and understanding. In addition, interviews and ethnographic studies that are digitized and coded can be used for thematic and discourse analysis concerning terminology and the process of understanding, as can the compilation and analysis of social media content.

The framework of the Sociocognitive Terminology theory of understanding can contribute in an interdisciplinary way (together with legal studies, language sciences, Computer Science, Digital Humanities technologies and resources, cognitive sciences) to the study of the effects of A.I. on multilingualism, in order to promote practices that respect linguistic rights which could contribute to the promotion of multilingualism and to the study of linguistic equality and variations of languages.

Frame-based Terminology

Pamela Faber

University of Granada

This chapter develops the theoretical premises and the semantic, syntactic, and pragmatic microtheories underlying Frame-Based Terminology and the organization of specialized knowledge. It also explains how these premises have been implemented to formulate terminological definitions, model specialized knowledge concepts, create semantic networks, and design specialized knowledge resources. Conceptual modeling in knowledge resource design should capture both the micro- and macrocontexts of concepts. This not only means structuring individual term entries but also capturing the relationships between them. These relations reveal the most frequent combinations and activations of specialized knowledge units, which are indicative of some type of large-scale knowledge structure in which individual concepts are meaningfully related to each other.

Keywords: Frame-Based Terminology, knowledge representation, conceptual modeling, knowledge resource design, terminological definition, semantic network

1. Introduction

One of the most crucial issues that must be addressed in any theory of Terminology is how specialized concepts should be represented so as to help users understand their meaning and their relations with other concepts. This type of knowledge representation or conceptual modeling should capture and account for both the micro- and macrocontexts of concepts. It signifies not only representing individual concepts and their terminological designations, but also integrating them into larger knowledge structures or frames in order to map the relations that they hold with others. This is the basis for Frame-based Terminology (FBT).

FBT is a cognitive approach to Terminology, which shares many of the same premises as the Communicative Theory of Terminology (CTT) (Cabr  2003) and the Socio-Cognitive Theory of Terminology (SCTT) (Temmerman 2000, this volume). It resembles these theories in that it is also a descriptive and text-driven

approach, which admits term variation as well as polysemy. However, Frame-based Terminology specifically focuses on specialized knowledge representation, category organization and description, as well as the semantic and syntactic behavior of terminological units in one or various languages.

FBT also maintains that knowledge of conceptualization processes, as well as the organization of semantic information in the brain, should underlie theoretical assumptions concerning the access, retrieval, and acquisition of specialized knowledge as well as the design of specialized knowledge resources (Faber 2011; Faber, Verdejo, et al. 2014).

As shall be explained, frames can be used to link terms in different languages to the same specialized concept. Conceptual structure should thus be made explicit in term entries so that users can locate each concept within a specialized domain and view its relations with other concepts. From a linguistic perspective, these semantic relations should be reflected in the terminological definition, which is the natural language representation of conceptual structure.

Frame-based Terminology is composed of a set of microtheories: (i) a semantic microtheory focusing on term meaning, definitions, and conceptual organization; (ii) a syntactic microtheory that analyzes specialized phraseology and the structure of multi-word terms; and (iii) a pragmatic microtheory that explains the cultural and contextual parameters of specialized communication (Faber 2015). Each microtheory is related to the information in term entries, the relations between specialized knowledge units, and the concepts that they designate.

The rest of this paper is structured as follows. Section 2 discusses the representation of frames in the brain, in general language, and in terminology; Section 3 outlines and explains the different microtheories in FBT. Section 4 lists the conclusions and examines future challenges.

2. Frames in Frame-based Terminology (FBT)

FBT applies the notion of ‘frame’ (Minsky 1975; Fillmore 1985) defined as a schema or knowledge structure, which relates elements and entities associated with a particular scene, situation that is part of human experience. As humans, we do not store verbatim copies of experiences in our memory. Rather, we integrate new incoming information from the surroundings in relation to our pre-existing knowledge about the world. This knowledge is accumulated across ontogenetic development through experiences during which the individual detects regularities in the environment (Brod et al. 2013). These regularities are the basis for frames. A frame is thus an organized package of knowledge that humans retrieve from long-term memory to make sense of the world (Faber 2012). In this sense, it is a kind of cognitive context,

which has a parallel in language in the sense that an entire frame of related meanings can be activated by a single term.

For example, the term, *biodegradable*, activates a frame of positive environmental action. Under the right conditions, a *biodegradable* substance will eventually break down to its basic components and blend back in with the earth. This term thus elicits a scenario with various actors: (i) the substance (e.g. detergent or plastic) that is biodegraded or broken down; (ii) a process (biodegradation); and (iii) an agent (microorganisms, bacteria, fungi) that triggers the biodegradation process and removes a polluting substance from the planet. *Biodegradable* also alludes to the positive short-term and long-term effects of this process, such as a cleaner environment, environmental sustainability, enhanced ecofriendliness etc.

In FBT, a term is defined as the natural language designation of an object, process, action, or attribute within a specialized knowledge domain and its conceptualization. Used for purposes of specialist communication and knowledge acquisition, a term is a relational unit that triggers a rich knowledge representation in which visual, auditory, and linguistic data coalesce and converge in a frame, which takes the form of a semantic network of related concepts.

FBT assumes that a theory of Terminology should aspire to neurological and psychological adequacy. In fact, recent research in Neuroscience has found neurological correlates for frames in language (see Section 2.1). For this reason, knowledge of linguistic and conceptualization processes as well as the organization of semantic information in the brain should be considered and even underlie any theoretical assumptions concerning the representation of specialized concepts as well as the design of knowledge resources (Faber 2011; Faber and León-Araúz 2014, 2019).

2.1 Frames in the mind and brain

Although there is a general consensus that frames exist in the mind (Barsalou 1992, 2017), the nature of frames in the brain is controversial. In particular, there is no agreement as to whether and to what degree, sensorimotor information participates in semantic representation and processing (Kiefer and Pulvermüller 2012). Whereas some researchers maintain that concepts are stored as abstract symbols in a common format (Quillian 1969; Anderson 1983), others say that categorization and conceptual representations are dynamic and are largely based on sensorimotor information (Gallese and Lakoff 2005). However, reality probably lies somewhere in-between these two extremes. According to Lambon Ralph et al. (2017), frame-like configurations of knowledge reside in a brain network that connects the sensorimotor regions of the cortex to a semantic hub, where everything comes together and is translated into a common format (the neural correlate of a frame).

This theory, which is known as the 'hub-and-spokes' theory, seems to combine the best of both worlds (Patterson and Lambon Ralph 2016).

One of the premises of FBT is that semantic representation in the brain is relevant to the configuration, and representation of term entries in a knowledge base. In addition, the types of information activated by subject field experts also should have an impact on conceptual networks and term definitions. For example, in a pilot fMRI study, the brain activation images of expert geologists were compared to those of novices as they performed a series of different tasks, such as linking geological tools to their function and the geological tools to images (Faber, Verdejo, et al. 2014). The results of this study indicated that expert knowledge has a conceptual representation that transcends sensory input modalities (i.e. vision and hearing). This highlights the importance of context in the form of episodic memory of the interaction, location, and manipulation of objects such as scientific instruments. This type of context is the overarching scenario that subsumes the formation of contextual associations and ultimately of templates, as posited in FBT (Faber, Mairal, and Magaña 2011). This supports the creation of rich term entries in the form of specialized knowledge representations that combine visual, auditory, and linguistic data. This information converges in a frame in the form of a semantic network.

2.2 Frames in general language

According to Langacker (1987), language structure reflects conceptual structure. This signifies that human thought and thought processes can be analyzed by using language as a mirror of the mind. Just as language and thought are irrevocably intertwined, words cannot be separated from the concepts that they represent. Accordingly, if frames exist in the mind and brain and are related to semantic comprehension, it is reasonable to assume that they can also be found in language and are reflected in texts. This is the foundational premise of Frame Semantics (Fillmore 1982), which posits that all concepts are part of a larger structure (semantic frame), and are related in such a way that the activation of one lexical unit evokes the entire frame.

According to Barsalou (1992), frames constitute the general format of concept representation, and hence of categorization, in human cognition. He claims that frames can be used for modeling the structure of concepts as well as conceptualizations or the ad-hoc modifications of concepts in actual contexts. Frames have the advantage of being flexible since they can be reduced or enriched by removing or adding attributes, or by changing the specification of attribute values.

Frame Semantics has expanded its scope and now refers to a wide variety of approaches to the systematic description of natural language meanings. Generally speaking, Frame Semantics relates linguistic utterances to our conceptualization

of world knowledge, such as event types and their participants. A semantic frame thus refers to a collection of facts or a coherent structure of related concepts that specifies attributes, functions, interactions, etc., typically associated with a certain word or term. This process includes the integration of non-linguistic information and is the key to knowledge representation.

2.3 Frames in Terminology

Given the relevance of frames in general language, it is also reasonable to assume that frames are relevant to specialized knowledge. According to Schurz and Votsis (2015), frames play a central role in scientific thinking and can even explain scientific theory change. For example, correspondence relations between frame representations of the phlogiston theory and Lavoisier's oxygen theory illustrate why phlogiston theory, despite being erroneous, was so popular over a long period of time.

From a more practical perspective, knowledge representation is central to Terminology since terminological work, for whatever purpose, involves modeling specialized knowledge concepts within a given specialized domain so that users can understand them better. Conceptual modeling in knowledge resource design involves capturing both the micro- and macro-contexts of concepts.

Despite the importance of frames, situations, or events in cognitive processing, contextual information is rarely found in specialized knowledge resource, much less systematically represented. According to Bowker (2011), term banks generally present terms without any specification of their linguistic context, possibly because of the belief that terms in the same specialized field are monosemic. However, this is a questionable assumption because concepts can be recategorized so as to constrain their relational behavior. Terms can thus have variants with different projections (Faber and León-Araúz 2016). A crucial aspect of knowledge base design is thus to provide a relational view of concepts within a larger knowledge configuration or *frame*.

As previously mentioned, a frame in FBT (Faber 2012, 2015) is an organized package of knowledge retrieved from long-term memory to make sense of the world. In its simplest form, a frame is an underlying conceptual structure into which the meanings of a number of related terms fit. FBT makes Busse's (2012) distinction between predicative frames and concept frames. Predicative frames are used for the description of verbs and their nominalizations. They represent events and states of affairs in terms of their situation types and participants. Concept frames mostly represent entities designated by nouns and noun phrases. They are what Barsalou (1992) refers to when he states that frames are a general format for the representation of categories.

As explained in Section 3, FBT frames are extracted from terminological definitions as well as from corpus texts in different languages through the use of knowledge patterns that encode semantic relations (Meyer 2001; Marshman 2007a, chapter in this volume). The data thus obtained are used to structure categories to create frames for individual concepts as well as to characterize general processes and actions. Although definitions and corpus data are used to extract information, the assumption is that the resulting frames encode conceptual knowledge that is non-language-specific. This is based on the premise that both the semantics and syntax of language can be regarded as a mirror of conceptual structure. Frames thus primarily stem from the nature of the concept itself as well as its combinatorial potential.

3. Microtheories in Frame-based Terminology

Frames and their components in FBT are derived from language, namely from term definitions and corpora. This means that the methodology for frame construction is based on linguistic analysis. Since it is difficult to develop a single comprehensive theory to account for this type of analysis, a realistic alternative is to develop a set of microtheories responsible for manageable-size chunks of it (Nirenburg and Raskin 2004, 29). The specification of the behavior of specialized language units in FBT is thus based on the following set of microtheories: (i) a semantic microtheory; (ii) a syntactic microtheory and (iii) a pragmatic microtheory.

3.1 Semantic microtheory

According to Cognitive Semantics (Talmy 2000), lexical meaning is a manifestation of conceptual structure. Both general and specialized lexical items can be regarded as conceptual categories of distinct yet related meanings that exhibit typicality effects. In this regard, conceptual modeling can benefit from the semantic analysis of linguistic concepts.

In specialized language, semantics is mainly concerned with the representation of terminological units and their relations with other units in the same domain. Basic areas of focus include the following: (i) the classification of specialized language units (and the concepts thus designated) into concept classes; (ii) concept description; (iii) the specification of conceptual relations.

3.1.1 *Classification of concepts*

Categorization involves classifying concepts in categories, based on shared characteristics. This is usually done for a specific purpose and is a fundamental process in cognition and language. The organization of conceptual space is crucial in Terminology, which focuses on the mapping of specialized knowledge domains. This involves establishing the major conceptual categories in the subject field, ascertaining which concepts qualify for membership in each category, and specifying their relations with each other.

Category membership and disjunction are guided by a set of conceptual characteristics and subdivision criteria. The reason for grouping entities together may be because they have the same function or purpose, similar physical, spatial or temporal characteristics, the same role in a structure of other entities or concepts, or some combination of these or other factors (Michalski 1991, 4).

To specify and classify the concepts in a specialized knowledge field, it is usual to begin with a set of terms that are regarded as the most characteristic of that domain. These can be extracted from a specialized knowledge resource of prestige or from a domain-specific corpus of texts, based on their frequency or salience. Depending on their meaning, the terms are first grouped in ontological starter categories such as ENTITY, PROCESS, ATTRIBUTE, etc. (Mahesh and Nirenburg 1995; Moreno-Ortiz and Pérez-Hernández 2002). Because of their dynamism (Kageura 2002), certain concepts may be members of more than one category.

For example, physical concepts are those occupying space and time. They may be natural or artificial objects or processes. Abstract concepts include theories, units of measurement, equations, etc. (Faber, León-Araúz, and Prieto Velasco 2009). Such criteria are crucial in terminology work, where the focus is on what differentiates concepts from each other (Madsen and Thomsen 2009).

The semantic categorization of the concepts in EcoLexicon (ecolexicon.ugr.es), the practical application of Frame-based Terminology, was carried out by determining degrees of specificity (Murphy and Lassaline 1997) and conceptual similarity (Hahn, and Chater 1997; Faber, León-Araúz, and Reimerink 2014). Each concept was tagged with a category showing its interrelation with ontologically similar elements (Gil-Berrozpe, León-Araúz, and Faber 2019). These semantic classes were mainly based on concept definitions and the contextual information in the EcoLexicon corpus. They were validated by contrasting them with the ontological classes in ENVO (Buttigieg et al. 2013, 2016). As described in Gil-Berrozpe, León-Araúz and Faber (2019), the 4,500 concepts in EcoLexicon were classified in 152 categories, distributed in five categorization levels. The most general level is composed of the previously mentioned three starter ontological categories:

- A: ATTRIBUTE – properties of entities and processes
- E: ENTITY – physical and mental objects
- P: PROCESS – events extending over time and involving different participant

Depending on their ontological nature, concepts were subclassified in up to five levels of specificity, as exemplified in the following hierarchy of CREATION concepts in Table 1 (Gil-Berrozpe, León-Araúz, and Faber 2019).

Table 1. Hierarchy of CREATION concepts

E: ENTITY	
E-1:	CREATION
E-1.1:	ARTIFACT (e.g. <i>dc bus</i>)
E-1.1.1:	CONDUIT (e.g. <i>duct</i>)
E-1.1.2:	CONTAINER (e.g. <i>sedimentation tank</i>)
E-1.1.3:	INSTRUMENT (e.g. <i>centrifugal pump</i>)
E-1.1.3.1:	MEASURING INSTRUMENT (e.g. <i>accelerometer</i>)
E-1.1.3.2:	RECORDING INSTRUMENT (e.g. <i>albedograph</i>)
E-1.1.3.3:	SAMPLING INSTRUMENT (e.g. <i>automatic sampler</i>)
E-1.1.3.4:	TRANSFORMING INSTRUMENT (e.g. <i>solar cell</i>)
E-1.1.4:	VEHICLE (e.g. <i>dredger</i>)
E-1.2:	SOFTWARE (e.g. <i>computer application</i>)
E-1.3:	STRUCTURE (e.g. <i>pier</i>)
E-1.3.1:	BUILDING (e.g. <i>oil refinery</i>)
E-1.3.2:	DEFENSE STRUCTURE (e.g. <i>reef breakwater</i>)

Additionally, multidimensional concepts were classified in as many hierarchies as necessary, depending on the salient features in their definitions and in the corpus. For instance, as can be observed in Table 2, one of the most multifaceted concepts is PORT, which is a member of four categories (Gil-Berrozpe, León-Araúz, and Faber 2019).

Table 2. Conceptual categorization of PORT

Concept	PORT
Definition	natural or artificial place along a river or sea coast that gives ships and boats protection from storms and rough water, and where ships can load and unload cargo.
Conceptual category	E-1.3: STRUCTURE E-4.1: ARTIFICIAL GEOGRAPHIC FEATURE E-4.2: NATURAL GEOGRAPHIC FEATURE E-12.1.2: FACILITY

A port can thus be regarded as a structure because it is composed of a set of buildings. It is an artificial geographic feature because it is a location that offers vessel protection. When a port is a safe harbor because of its geographical characteristics, it is a natural geographic feature. A port can also be considered a facility when it has the specific purpose of unloading cargo.

3.1.2 Concept description

Specialized concepts are designated by terms. Similarly, each concept is described by a definition, which is a natural language explanation of its location in the conceptual structure of the specialized domain. A definition not only specifies the properties of a concept, but also links it to others (Antia 2000, 113–115).

A definition should thus be coherent with the concept itself, its semantic category, and the set of concepts related to it. In fact, it should provide a blueprint of the concept as well as its entire context. Maximally coherent definitions should follow the pattern or template typical of the semantic category. For purposes of information retrieval, the same phrases or knowledge patterns should be used to encode semantic relations.

Most definitions of specialized knowledge units are composed of a generic or superordinate term and differentiating features (Eck and Meyer 1995, 83–87; Sager 1990, 42), reflecting an external semantic representation, which relates other concepts to the concept being defined. The resulting network is composed of hierarchical relations, such as *type_of* and *part_of*, but also includes non-hierarchical semantic relations (e.g. *affects*, *result_of*, *causes*, *located_at*, etc.) (Faber, Mairal and Magaña 2011, 568).

Frame-based Terminology (FBT) maintains that each conceptual category has a template that can be used to model the definitions of category members. The set of semantic relations in this template depends on the nature of the category. For example, basic characteristics underlying the representation of a man-made object or artifact such as a SCIENTIFIC INSTRUMENT include its subtype, parts, function, and context of manipulation (Table 3).

Table 3. INSTRUMENT category template (Faber and León-Araúz 2014)

INSTRUMENT	
<i>type_of</i>	Type of instrument
<i>has_parts</i>	Parts of the instrument
<i>has_function</i>	Function of the instrument
<i>has_agent</i>	Agent that manipulates the instrument

For example, a RECORDING INSTRUMENT (MARIGRAPH, PLUVIOGRAPH, ANEMOGRAPH, etc.) is a subtype of INSTRUMENT. As an artifact, a recording instrument has a function (i.e. RECORDING) as well as a recorded phenomenon (TIDES, RAIN, WIND). As a tool, it is operated by humans and thus activates a simulation frame in which much of the perceiver's knowledge of the artifact involves his/her ability to manipulate it, place it at a certain location, and in some way, extract information from it.

3.1.3 *Specifying semantic relations*

Another major issue in the design of any knowledge base is to link the concepts in a domain by means of semantic relations. The inventory of relations affects the informativity of the terminological resource. When semantic relations in term bases are restricted to *type_of* and *part_of* relations, this is conducive to static configurations, which are at odds with the need to represent dynamic action in domain models (Barrière 2001, 137). According to Rogers (2004, 218–219), there is a vast array of ontological relations available to enrich concept systems.

Terminological knowledge bases can be more coherent and dynamic when the semantic relations contemplated also encompass horizontal or associative relations. Such relations largely depend on the type of entity described, its nature, and relational power. The top-level identity of the concept influences its core inventory of relations. Accordingly, these relations not only depend on the nature of the concept, but also on the nature of the domain. Furthermore, the key set of semantic relations differs from subject field to subject field.

As previously mentioned, the category templates in Frame-based Terminology consist of a set of relations typical of the category, some of which are domain-specific. The semantic relations in EcoLexicon (ecolexicon.ugr.es) are the following: *type_of*, *part_of*, *made_of*, *phase_of*, *delimited_by*, *located_at*, *attribute_of*, *result_of*, *affects*, *causes*, *has_function*, and *effected_by*. In this inventory, *type_of* and *part_of* are vertical relations. The relations, *made_of*, *phase_of*, *delimited_by*, *located_at*, and *attribute_of*, can be regarded as subtypes of meronymy. This distinction is in consonance with Winston, Chaffin, and Herrmann (1987) because not all parts interact in the same way with their wholes. The six different meronymic relations are not only based on domain-specific needs, but especially on ontological reasoning and consistency (León-Araúz 2009).

The semantic relations, *affects*, *causes*, *has_function*, and *effected_by* are horizontal relations. *Has_function* has three domain-specific subtypes: *measures*, *studies*, and *represents*. Evidently, all semantic relations in the inventory have a corresponding inverse relation. Examples of these relations are shown in Table 4.

Table 4. Semantic relation types in EcoLexicon (ecolexicon.ugr.es)

Semantic relations	Concept 1	Concept 2	Examples
<i>type_of</i>	PHYSICAL ENTITY	PHYSICAL ENTITY	masonry dam <i>type_of</i> dam
<i>part_of</i>	PHYSICAL ENTITY	PHYSICAL ENTITY	main layer <i>part_of</i> breakwater
	MENTAL ENTITY	MENTAL ENTITY	microbiology <i>part_of</i> biology
<i>phase_of</i>	PROCESS	PROCESS	pumping <i>phase_of</i> dredging
<i>made_of</i>	PHYSICAL ENTITY	PHYSICAL ENTITY	air <i>made_of</i> gas
<i>located_at</i>	PHYSICAL ENTITY	PHYSICAL ENTITY	jetty <i>located_at</i> canal
<i>takes_place_at</i>	PROCESS	PROCESS	littoral transport <i>takes_place_at</i> sea
<i>delimited_by</i>	PHYSICAL ENTITY	PHYSICAL ENTITY	stratosphere <i>delimited_by</i> stratopause
<i>result_of</i>	PROCESS	PROCESS	aggradation <i>result_of</i> sedimentation
<i>causes</i>	PHYSICAL ENTITY	PROCESS	water <i>causes</i> erosion
<i>affects</i>	PHYSICAL ENTITY	PROCESS	groyne <i>affects</i> littoral transport
	MENTAL ENTITY	ENTITY	pesticide <i>affects</i> water
	PHYSICAL ENTITY	ENTITY	wave <i>affects</i> groyne
	PROCESS	PROCESS	precipitation <i>affects</i> erosion
<i>has_function</i>	ENTITY	PROCESS	aquifer <i>has_function</i> human supply
<i>attribute_of</i>	PROPERTY	ENTITY	abyssal <i>attribute_of</i> plain
	PROPERTY	PROCESS	anthropic <i>attribute_of</i> process

3.1.3.1 Automatic retrieval

Given the importance of semantic relations, the automatic retrieval of related term pairs has been explored from different perspectives. One of the most useful extraction methods is based on knowledge patterns (KPs) (Barrière 2004; Condamines 2002; Lafourcade and Ramadier 2016). The term KP refers to a lexico-syntactic pattern between terms encoded in a proposition in real texts.

Examples of hyponymic knowledge patterns are the following: *X is a kind of Y*, *comprise(s)*, *consist(s)*, *define(s)*, *denote(s)*, *designate(s)*, *is/are*, *is/are called*, *is/are defined as*, *is/are known as*, etc. (R. Meyer 1993; Pearson 1998). Typical meronymic KPs include the following patterns: *part of*, *constituent of*, *constituted by*, *made of*, *composed of*, *contains*, etc. (Berl and Charniak 1999; Girju, Badulescu and Moldovan 2003).

These relations may be codified by prepositional phrases, possessives, and partitive verbs, though many KPs can also be polysemic. For instance, *including* expresses both hyponymy and meronymy; and *formed by* expresses meronymy and causality (León-Araúz 2014).

In the case of KPs for non-hierarchical relations, causation can be expressed by passive, active, subject-object, nominal, and verbal propositions. Moreover, causation is not only expressed by constructions such as *due to* or *because of*, but also by causative nouns (*cause* or *consequence*) and verbs. Although there are many causative verbs (e.g., *cause*, *generate*, *lead*, *produce*, etc.), their syntactic behavior can also vary (León-Araúz, San Martín and Faber 2016). Typical knowledge patterns used in Frame-based Terminology are shown in Table 5.

Table 5. Knowledge patterns in León-Araúz and Reimerink (2010, 8)

Semantic relation	Knowledge pattern
<i>is_a</i>	such as, rang* from, includ*
<i>part_of</i>	includ*, consist* of, formed by/of
<i>made_of</i>	consist* of, built of/from, constructed of, formed by/of/from
<i>located_at</i>	form* in/at/on, found in/at/on, tak* place in/at, located in/at
<i>result_of</i>	caused by, leading to, derived from, formed when/by/from
<i>has_function</i>	designed for/to, built to/for, purpose is to, used to/for
<i>effected_by</i>	carried out with, by using

3.1.3.2 Knowledge-rich contexts and sketch grammars

Studies on FBT knowledge patterns such as Rojas-García and Cabezas-García (2019) explain how semantic relations can be extracted from corpus texts in different languages by searching for knowledge-rich contexts (KRCs).

A KRC has at least one item of domain knowledge useful for conceptual analysis (I. Meyer 2001, 281). However, high-density KRCs, such as the following, contain different semantic relations (e.g., hyponymy, meronymy, and causality) that reflect links between many different concepts (León-Araúz and Reimerink 2019, 110).

1. Nuées ardents *are but one type of* [type_of] *of* pyroclastic flow, *which include a variety of mixtures of* [has_a] volcanic blocks, ash, gas, and lapilli *that produce* [has_result] volcanic rocks *called* ignimbrites.
2. Contact metamorphism *of* [attribute_of] carbonate rocks *produces* [has_result] minerals *such as* [has_type] wollastonite, tremolite and grossular garnet, spessartine garnet and andradite garnet.

In FBT, knowledge pattern queries in the form of micro-grammars (León-Araúz and Reimerink 2018; León-Araúz and San Martín 2018; León-Araúz, San Martín and Faber 2016) are used to extract concordance lines for a term and analyze its combinatorial potential. EcoLexicon Semantic Sketch Grammars (ESSGs) are based on corpus queries in Sketch Engine (Jakubíček et al. 2010), which formalize syntactic

patterns in the form of regular expressions combined with POS tags. ESSGs can be used to extract semantic relations and provide semantic word sketches. In fact, users can even apply them to their own corpora

Sketch grammars are collections of corpus query language (CQL) expressions that allow users to automatically produce word sketches. The ESSG extracts semantic word sketches based on some of the most common semantic relations in terminology: generic-specific, part-whole, location, cause, and function.

In domain-specific specialized language, some patterns always have the same form and order (e.g. *such as*), whereas others differ, depending on the directionality of the pattern. The same sentence may also produce more than one term pair because of the enumerations on each side of the pattern (e.g. *x, y, z and other types of w*) (León-Araúz and San Martín 2018). Table 6 shows a simplified version of the patterns included for the generic-specific, part-whole, and cause relations.

Table 6. Simplified version of the patterns in each ESSG
(León-Araúz and San Martín 2018, 95)

Generic-specific	HYPERNYM, ([: is belongs (to) (a the ...) type category ... of HYPERNYM // types kinds ... of HYPERNYM include are HYPERNYM // types kinds ... of HYPERNYM range from (...) (to) HYPONYM // HYPERNYM (type category ...) (, () ranging (...) (to) HYPONYM // HYPERNYM types categories ... include HYPONYM
Part-whole	WHOLE is comprised composed constituted (in part) of by PART // WHOLE comprises PART // PART composes WHOLE // PART is constitutes (a the ...) part component ... of WHOLE // WHOLE has includes possesses (...) part component ... (, () (: such as usually namely ...) PART // WHOLE has includes possesses (a the ...) fraction amount percent... of PART
Cause	CAUSE (is) responsible for EFFECT // CAUSE causes produces ... EFFECT // CAUSE leads contributes gives (rise) to EFFECT // CAUSE-driven -induced -caused EFFECT // EFFECT (is) caused produced ... by because due (of to) CAUSE

Sketch grammar rules are established by first collecting KPs and translating them into CQL sketch grammar rules. These rules are then enriched and refined to balance noise and silence. They are then evaluated with a precision and recall analysis, after which new patterns and modifications may be included (León-Araúz and San Martín 2018, 96).

Evidently, no type of information extraction would be possible without a corpus. One of the products of FBT is the EcoLexicon Corpus, initially created as an internal tool for knowledge extraction.

3.1.3.3 *The EcoLexicon corpus*

The EcoLexicon English Corpus (EEC) is a 23.1-million-word specialized corpus of contemporary environmental texts. It was compiled by the LexiCon research group for the development of EcoLexicon (ecolexicon.ugr.es) (Faber, León-Araúz and Reimerink 2016; San Martín et al. 2017). The EEC was processed with an internal application of the research group. Then it was recompiled within Sketch Engine with the Penn Treebank tagset (TreeTagger version 3.3) and with the EcoLexicon Semantic Sketch Grammar (ESSG) (León-Araúz and San Martín 2018; León-Araúz, San Martín and Faber 2016).

In EcoLexicon, the EEC and its Spanish counterpart (over 70 million words) can be queried with pragmatic restrictions such as author, date of publication, target reader, contextual domain, and keywords (León-Araúz, San Martín and Reimerink 2018). Each text in the EEC is tagged with the following XML-based metadata: (i) domain; (ii) user; (iii) geographical variant; (iv) genre; (v) publisher; (vi) year; (vii) country. In the future, the EEC will be annotated with semantic tags to improve its querying potential. These tags will include semantic categories and argument structure.

3.2 Syntactic microtheory

The syntax of specialized knowledge units has become increasingly relevant to Terminology. This is partly due to the fact that syntactic patterns are better described by specifying the lexical classes to which they apply (Geeraerts 2015). This interface between syntax and semantics is especially relevant in Terminology, which is based on the conceptualization and categorization of specialized concepts designated by terms.

In this regard, terms have a combinatorial value and distinctive syntactic projections, depending on their semantic category (Cabezas-García and Gil-Berrozpe 2018; Cabezas-García and León-Araúz 2018). This is reflected in phraseological patterns in specialized texts as well as in multi-word terms (MWTs) (Horsella and Pérez 1991; Daille, Dufour-Kowalski and Morin 2004; Hendrickx et al. 2013). Since MWTs consist of two or more juxtaposed concepts (Zelinsky-Wibbelt 2012), understanding their meanings depends on the user's ability to successfully retrieve the relation between concepts.

MWTs are composed of a head noun modified by one or more elements, usually nouns or adjectives (Levi 1978) (e.g. *dune erosion*, *significant wave height*). As such, they reflect the preference of Germanic languages for semantically condensed structures, in which economy of expression prevails over clarity of expression (Štekauer et al. 2012; Sanz-Vicente 2012).

In English, the most frequent pattern is pre-modification in which one or more modifiers are stacked on the left of the head (e.g. *wind power*, *wind power generation*, *wind power generation system*). However, post-modification is also possible (*angle of internal friction*), often combined with one or various premodifiers (e.g., *effective angle of internal friction*) (Cabezas-García and Gil-Berrozpe 2018).

Some MWTs are formed by predicate deletion or predicate nominalization (Levi 1978). They can thus be analyzed in terms of their underlying propositions. In MWTs formed by predicate deletion, modifiers are usually the direct object of the predicate, which has been elided (e.g. *power plant* derived from ‘a plant produces power’).

Moreover, in MWTs formed by predicate nominalization, the head is a nominalized verb, whose modifiers are the subject of the predicate (*oil pollution* derived from ‘oil pollutes X’) or the direct object of the predicate (e.g. *air pollution* derived from ‘X pollutes air’). Representing and understanding the meaning of MWTs thus entails the specification of the underlying predicate and its arguments.

Predicate-argument structure refers to the lexical representation of argument-taking lexical items (Levin 2013). These are typically verbs and their nominalizations. The specification of argument structure involves identifying the number of arguments that a predicate can take, their syntactic expression, and their semantic relation to the predicate. The concealed or nominalized verbs and their arguments are thus crucial to understanding the meaning of MWTs (Cabezas-García and Faber 2017c).

In FBT, the relation of a predicate to its arguments and adjuncts (Tesnière 1959) is referred to as a *microcontext*. MWTs designate the convergence of two or more concepts. However, the combination of concepts is not random but rather is based on semantic constraints. Cabezas-García (2019, 281) found that the concepts (as reflected in the terms that designate them) that belong to a certain category have similar combinatorial patterns. In other words, the semantic categories that generally act as the head of the MWT open a set of slots, which are in direct relation to their meaning. These slots determine the formation of MWTs that further specify the head.

Since the syntactic-semantic core of an MWT is its underlying predicate, MWT analysis is based on argument structure (Cabezas-García and Faber 2017a, b, c, 2018). In FBT, MWTs are represented and disambiguated by means of semantic roles (Agent, Patient, Theme, Result, etc.). For example, *pollution* is the nominalization of *pollute*, a verb that can have an *agent* (polluting entity), and/or a patient (polluted entity). This means that even though *oil pollution* and *air pollution* have the same syntactic structure, their semantic structure is quite different. More specifically, *oil pollution*, *oil* has the semantic role of Agent since it is the polluting substance. In contrast, in *air pollution*, *air* is the Patient because it is the polluted substance.

However, there are other complementary ways of expressing this meaning. According to Nakov and Hearst (2006), the meaning of an MWT can also be represented by means of a verb paraphrase or a set of paraphrases (Teubert 2005). For example, in *beach erosion*, *beach* is affected by *erosion*, which means that erosion *takes place at/occurs/affects/threatens/impacts* the beach.

In FBT, this type of analysis is further enriched when the use of paraphrases and KPs is combined with the semantic tagging of the constituents of the MWT. These tags, which specify semantic categories, help to disambiguate and infer the semantic relation between the constituents of the MWT (Cabezas-García and León-Araúz 2018). This means that the combinations of concepts in MWTs are not random, but rather the result of semantic constraints (Štekauer 1998; Cabezas-García and Gil-Berrozpe 2018).

The basic assumption is that the head of the MWT has slots (Maguire, Wisniewski and Storms 2010), which are filled by entities belonging to a certain semantic category, each of which fulfills a certain role. For example, *emission*, defined as “release of a substance into the environment as a result of a process” has three slots: (i) SUBSTANCE; (ii) SPACE; (iii) PROCESS.

In MWTs for *emission*, the semantic category of SUBSTANCE has the role of THEME since it refers to the entity being emitted (i.e. *greenhouse gases* and *CO2*). The second slot, which represents the category of SPACE, refers to the environment in which substances are emitted (i.e. *air emission*, *atmospheric emission*). The combination of both slots is also possible (*atmospheric emission of CO2*). The semantics of the head thus determines possible modifiers, and thus the generation of hyponymic MWTs, whose components further specify the semantic slots in the meaning of the head. This type of analysis reflects the interface between syntax and semantics and explains the systematic generation of MWTs, based on their microcontext.

3.3 Pragmatic microtheory

In FBT, the pragmatic microtheory accounts for context-dependent aspects of specialized meaning. It seeks to characterize the features of the communicative context that help to determine which proposition is expressed by a term or terminological phraseme. The meaning of a term is thus regarded as a function of its context, whose scope can be local or global (Faber and León-Araúz 2016). In FBT, context types are linguistic, cultural, and graphical.

3.3.1 *Linguistic contexts*

Linguistic contexts link all the information within the term entry to the domain event. The most important linguistic context is the terminological definition, which describes the most general meaning applicable in the widest range of communicative

settings. In corpus texts, local linguistic contexts are usually spans of five items before and after the term occurrence. They are important for: (i) term disambiguation; (ii) meaning definition formulation; (iii) specification of linguistic usage; (iv) conceptual modeling; and (v) term extraction (Faber and León-Araúz 2016). Such contexts are based on syntagmatic patterns, and/or semantically mapped in terms of the interaction, foregrounding, or specification of the definitional features of the concepts activated in them.

In FBT, local linguistic contexts are reflected in concordance lines for each term. As previously mentioned, corpus concordances reflect Knowledge-Rich Contexts (KRCs). The entry point to the discovery of KRCs is the automatic identification of knowledge patterns (KPs), which indicate semantic relations (Barrière 2004).

3.3.2 *Cultural contexts*

Cultural contexts reflect the pragmatic value of terms, mainly in the form of parameters of terminological variation. The same specialized concept can often have many linguistic designations depending on the communicative context. Alternatively, the same linguistic designation can also refer to various concepts.

3.3.2.1 *Parameters of terminological variation*

Terminological variation is a cultural phenomenon since it is based on the user or usage context. User-based parameters are geographic, temporal or social variation, whereas usage-based parameters are field, tenor, and mode (Gregory and Carroll 1978). Nevertheless, there are other reasons for terminological variation. Freixa (2006, 52, this volume), for example, classifies the causes for terminological variation in the following categories: (i) dialectal, based on the origin of the authors; (ii) functional, based on communicative register; (iii) discursive, based on the style; (iv) interlinguistic, based on language contact; and (v) cognitive, based on different conceptualizations.

Since such variants often have a communicative and/or cognitive motivation, the use of a certain term may affect the semantics of a concept or the communicative context in which the concept is activated. Moreover, in those cases where the same concept can be designated by different dialectal variants stemming from the geographic origin of the author, this can also mean that the same term can be used to designate different entities in different cultures.

Cultural contexts are thus reflected in terms that are pointers to larger (global) situational, linguistic, and cultural contexts. Term variation is thus not isolated from conceptual and cultural representations since it is one of the manifestations of the dynamicity of categorization and expression of specialized knowledge (Fernández-Silva, Freixa and Cabré 2014).

3.3.2.2 *Cultural profiles or semplates*

The specification of a pan-European concept system for the languages in a specialized knowledge base requires a typology of cultural contexts or profiles linked to the most prominent semantic categories. Burenhult and Levinson (2008, 144) propose the term, *semplate*, which refers to the cultural themes or linguistic patterns that are imposed on the environment to create, coordinate, subcategorize, or contrast categories.

This is related to the supramodal representation with multiple levels of input, proposed by Binder and Desai (2011). The top level consists of schematic representations that are fleshed out by sensory-motor-affective input. These modality-invariant representations are similar to geographic maps (Lambon Ralph et al. 2010), in which each map type (geological, political, linguistic, etc.) codes the same chart/grid system, but differs in the presence or absence of each type of feature. No specialized knowledge concept can be activated in isolation, but rather is understood as part of a ‘situated’ event in which perception, culture, and a wide range of other dynamic factors come into play.

For example, the way that culture can shape conceptual representation is reflected in the concepts of WIND. Though invisible, wind is still as real as any geographic landform or feature. The concept of WIND has a cultural context, represented in terms of perception and action. Although its top-level schema constrains perceptual input, at the same time, it is also derived from sensorimotor mappings (Faber and Medina-Rull 2017). Through the ages, the Greeks, Romans, and Polynesians explained this phenomenon by creating a god or gods in charge of the wind. This supported the folk belief that the wind must come from somewhere and even have its own habitat.

From a meteorological perspective, winds are classified in terms of spatial scale, speed, direction, region of occurrence, and effect. However, even highly technical categorizations of wind types are mostly based on directionality and intensity. Precisely because the definition of wind depends on the perceivers and their orientation, it is evaluated by its effect on the landscape and on the inhabitants of the landscape. This is particularly the case of local winds, which are culture-bound entities that can affect choice of habitat, buildings, crops, and even vehicles for travel.

The semplate for the description of local wind is derived from the basic template for wind in general: *is_a*, *action_of*, *location_of*, and *direction_of*. As small-scale winds arising from differences in temperature and pressure in localized areas, local winds are more specific and thus have additional relations that characterize them, namely, *temperature_of*, *time_of*, *duration_of*, *result_of*, *intensity_of*, and *water_content_of*. The parameters in Table 7 reflect how the inhabitants of a landscape understand the wind, and how they interact with it.

For example, when a wind receives a name, this is generally because it significantly affects the environment and alters daily life in some memorable (usually negative) way. The cultural salience of negativity is also reflected in the names of local winds. Although many of these terms lexicalize wind direction (*southeaster*, *nor'easter*, *nor'wester*) and location (*pampero*, *papagayo*), others foreground negative qualities.

For example, in regards to *action_of*, haboob is a strong wind in the Sudan that blows sand and dust. The relation *duration_of* is lexicalized in khamsin, which is derived from the Arabic *khamsun* or *hamsin* [fifty], the approximate number of days that this hot, dusty North African wind generally blows. The relation, *result_of* is negatively lexicalized in *wreckhouse winds* in Newfoundland.

Table 7. Semplate of LOCAL WIND (Faber and Medina-Rull 2017: 427)

Local wind	
General parameters	
<i>is_a</i>	Wind
<i>action_of</i>	Blow
<i>location_of</i>	geographic location where the local wind usually blows
<i>direction_of</i>	direction from which the local wind originates
Specific parameters	
<i>intensity_of</i>	strength of the wind
<i>time_of</i>	time of the year when the wind occurs
<i>duration_of</i>	temporal duration of the wind
<i>result_of</i>	effect caused by the wind
<i>temperature_of</i>	temperature of environment when local wind occurs
<i>water_content_of</i>	humidity of the wind

The name of a local wind can also embody malevolent supernatural entities, such as *diablo* [devil] a hot, dry, offshore wind in the San Francisco bay, or be conceptualized as a negative object, such as *simoom* from the Arabic *samma* [poison], a hot, dust-laden wind that blasts through the desert. This negative instrumental dimension of wind as an instrument causing pain or death is also reflected in the wind known as *el cordonazo* [the lash of St. Francis], a southerly hurricane wind along the west coast of Mexico.

Terms thus have an important cultural dimension that cannot be ignored, especially since certain conceptual categories are linked to the habitat, customs, and way of life of the speakers of a language and derive their meaning from the characteristics of the immediate environment.

3.4 Graphical contexts

The fourth type of context in Frame-based Terminology (FBT) is graphical. In EcoLexicon, the inclusion of different types of image enriches and enhances the linguistic information in other data fields and facilitates knowledge acquisition.

In FBT, images depict concepts and are selected so as to highlight their semantic relations. This means that the visual information in images reinforces and meshes with the information in terminographic definitions for a better understanding of concept systems (Faber, León-Araúz et al. 2007, 39). Accordingly, in FBT, iconic, abstract, or dynamic images in EcoLexicon (ecolexicon.ugr.es) are selected so that their features combine with conceptual propositions based on the following: (i) combinatorial potential of the concept; (ii) image types, such as photographs, drawings, and flow charts; (iii) morphological features or visual knowledge patterns (VKPs), such as labels, colors, arrows, and their effect on the functional nature of each image type (Reimerink, León-Araúz and Faber 2016).

VKPs are the means by which images convey conceptual knowledge. Reimerink and León-Araúz (2018) studied the interaction between concept type, image type, and VKPs, based on referential similarity and dynamism. Referential similarity refers to the degree to which an image resembles its referent in the real world on a scale ranging from non-similar to totally identical. Dynamism was also measured on a continuum ranging from totally static to very dynamic. The results showed which VKPs and which degrees of referential similarity and dynamism were most characteristic of a certain type of image and how they were related to the conceptual propositions represented.

VKPs, such as arrows, labels, and color-coding, were found to be polysemic since the same pattern could be used for different purposes in the same way that textual knowledge patterns were also able to convey different semantic relations (León-Araúz, Reimerink, and Faber 2019).

For example, an arrow, which is generally a dynamic VKP, does not transmit dynamism when used to connect a term to its representation in an image. However, when arrows describe a process, they generally convey dynamism and point in the direction of the next phase. The same is true for colors. In images with a high level of referential similarity, the colors in the image are the same or similar to those of the real world entity. In many cases, however, the function of the colors is not to realistically represent concepts or their natural surroundings, but rather to differentiate concepts that are closely related in time or space (Reimerink and León-Araúz 2018).

Although the conceptual knowledge underlying VKPs can only be interpreted in the context of each image, there is a certain combination of patterns, constrained by image and concept type, which make certain images more or less suitable for the representation of certain types of conceptual knowledge.

For example, León-Araúz and Reimerink (2018) evaluated the images in EcoLexicon and found that flow charts were more suitable for representing processes whereas photographs and drawings best depicted concrete entities and objects. Flow charts were also used to represent physical entities when they were defined as the agent or result of processes. Processes were represented by a sequence of drawings when they could be divided into phases and in photographs when the focus was on the result of the process.

Properties (e.g. SOIL PERMEABILITY) appeared more often in drawings because they were best conveyed by labels, which was the most frequent VKP in drawings. Not surprisingly, flow charts were also the image type that was used to convey procedural relations such as *result_of* and *causes*). In turn, photographs had a better representational adequacy for *type_of* and *has_location*. Drawings were used to represent all of the relations typical of physical entities: *type_of*, *has_location*, *part_of* and *delimited_by*. However, the *part_of* and *delimited_by* relations in all concepts were most often represented by drawings (León-Araúz and Reimerink 2018).

More research is necessary in order to gain further insights into the parameters that link visual contexts with linguistic contexts. However, based on research within FBT, it is evident that different types of image have an important role in specialized knowledge representation.

4. Conclusions

This chapter has outlined and explained the theoretical premises of Frame-Based Terminology (FBT) within the context of other communicative and cognitive theories of Terminology. Central to FBT is the concept of *frame*, a knowledge structure that situates the entities within a particular context that is part of human experience. One of the premises of FBT is that knowledge of linguistic and conceptualization processes as well as the organization of semantic information in the brain should underlie any theoretical assumptions concerning the retrieval and acquisition of specialized concepts as well as the design of knowledge resources. In this sense, recent research in neuroscience justifies rich term entries in the form of specialized knowledge representations that combine visual, auditory, and linguistic data, which would correspond to modality-specific sensory data. This information converges in a frame, which takes the form of a semantic network.

The semantic, syntactic, and pragmatic microtheories in Frame-based Terminology explain the premises on which the retrieval, specification, and analysis of specialized knowledge units is based. FBT can thus be regarded as a theory of contexts, based on knowledge representation in the brain and mind, as reflected in language.

The semantic microtheory explains how specialized knowledge concepts are classified by determining degrees of specificity and conceptual similarity. Concepts are described by definitions, which should follow the template that characterizes their respective semantic categories. Such templates consist of a set of vertical and horizontal relations typical of a given category, some of which may be domain-specific.

The syntactic microtheory highlights the importance of the interface between syntax and semantics in Terminology and specialized language. It focuses on the combinatorial value and distinctive syntactic projections of terms, stemming from their semantic category as well as their combinations with other categories. This is reflected in phraseological patterns in specialized texts as well as in multi-word terms (MWTs).

MWT analysis is based on argument structure as well as paraphrases. This analysis is also combined with the semantic tagging of MWTs, which helps to disambiguate and infer the semantic relation between constituents.

Finally, the pragmatic microtheory accounts for context-dependent aspects of specialized meaning. In FBT, such contexts can be linguistic, cultural, and graphical. Linguistic contexts are found in concordance lines, especially those that reflect Knowledge-Rich Contexts (KRCs). Their analysis is a key factor in the retrieval of semantic relations codified in knowledge patterns (KPs).

Cultural contexts reflect the pragmatic value of terms, mainly in the form of parameters of terminological variation. In addition to the standard user and usage-based parameters, these also include cultural profiles or *semplates*, which describe the way that culture and geographical location can shape conceptual representation.

Finally graphical contexts pertain to the different types of image used to enrich linguistic information and knowledge acquisition in a terminological knowledge base. Images can be iconic, abstract, or dynamic. Image selection should depend on the combinatorial potential of the concept, image type (photograph, flow chart, drawing, etc.), and visual knowledge patterns (VKPs).

In this regard, FBT is a framework for the multidimensional representation of specialized knowledge from a linguistic, cultural, and graphical perspective. Its premises facilitate specialized knowledge resource design, which should be able to capture and represent both the micro- and macrocontext of concepts. The main theoretical focus of FBT is thus on conceptual representation and on the semantic relations between concepts. Such relations reflect and reveal their most frequent combinations and activations, and are indicative of a larger-scale knowledge structure in which concepts become more meaningful.

Within FBT, there are currently many challenges, all of which are directly related to the three microtheories. For example, in regard to semantics, a major focus will be on the representation of interlingual and intralingual equivalence at the language and term levels (San Martín et al. 2020; León-Araúz, this volume). This would include a set of descriptive fields to be included at the term level and criteria according to which all terms can be grouped and compared (i.e., language level). Still another challenge in this area is the creation of domain-specific definitions adapted to user profiles (San Martín 2016). The syntactic microtheory should be able to account for the combinatorial patterns of single-word terms based on predicate-argument structure. This would also facilitate the prediction of multi-word terms and their automatic retrieval. Finally, the greatest challenge in pragmatics is the detection of culture-specific priorities and patterns in the designation of environmental entities. Without a doubt, the results obtained will provide valuable insights into terminology and specialized knowledge representation.

Funding

This research was carried out as part of project PID2020-118369GB-I00, *The Transversal Integration of Culture in an Environmental Knowledge Base*, funded by the Spanish Ministry of Science and Innovation.

Conceptual metaphors

José Manuel Ureña Gómez-Moreno and Francisco José Ruiz de Mendoza Ibáñez

University of Granada / University of La Rioja

This chapter offers an overview of conceptual metaphor research with particular attention to specialised language. It surveys the growing body of research showing the role of metaphor both as a tool to generate and develop scientific thinking and as a resource for specialised knowledge popularisation. This chapter also addresses specialised language studies that explore not only the embodied but also the socio-cultural dimension of metaphor in terminology, thus crucially refining conceptual metaphor theories and typologies in all spheres of communication.

Keywords: conceptual metaphor, Cognitive Linguistics, metaphor types, specialised language discourse, terminological figurative language, non-terminological figurative language

1. Introduction

The present chapter offers an overview of conceptual metaphor research with particular emphasis on specialised language. Section 2 gives an overview of conceptual metaphor research in general from a Cognitive Linguistics perspective. Recognised as a way to make sense of our experience, and thus central to a theory of knowledge (Lakoff 1987), metaphor has been approached from a lexical (Evans 2004), grammatical (Goldberg 1998) and pragmatic viewpoint (Sperber and Wilson 1986). Section 3 focuses on the significance of metaphor in specialised language and Terminology. Specifically, this second part surveys the growing body of research showing the pivotal role of metaphor both as a tool to generate, develop and communicate scientific and domain-specific thinking and as a resource for specialised knowledge popularisation. This section also addresses specialised language studies that explore not only the embodied but also the socio-cultural dimension of metaphor in terminology, thus crucially refining conceptual metaphor theories and typologies in all spheres of communication.

It should be highlighted that until the late 1970s, it was generally thought that metaphorical expressions and figurative language in general, had a marginal role in science. Indeed, Ortony (1993, 1) notes that, in the past, scientific and technological knowledge was regarded as precise and unambiguous. Hence, metaphor was believed to be unnecessary in scientific expression since meaning extensions lead to polysemy and, as a consequence, to misinterpretation and unscientific communication. Conceptual metaphor was later proven to have a direct bearing on the way (i) the human mind creates and categorises specialised concepts; and (ii) on the way such concepts are conveyed through language to structure and further all fields of expertise (cf. e.g., Alexiev 2005; Caballero 2006; Temmerman 2000; Ureña 2011; Ureña and Faber 2010).

For all these reasons, conceptual metaphor research has been highly instrumental and insightful in revealing fundamental aspects of human thought, language, and interaction that are central to scientific development and specialised language description. It can thus be concluded that conceptual metaphor theories – as will be shown – can be regarded as essential to terminology studies.

2. Conceptual metaphor: Emergence, underpinnings, and typologies

This section focuses on the emergence, development, classifications and applications of the notion of conceptual metaphor from the perspective of Cognitive Linguistics in its different manifestations. In line with the natural chronological evolution of conceptual metaphor, its description begins from a general linguistic perspective; however, this section also includes examples of terminological metaphor (figurative expressions from specialised language) to opportunely illustrate its relevance to the reader.

2.1 Cognitive Linguistics and conceptual metaphor

Over the past four decades, cognitive linguists (e.g., Lakoff 1993; Lakoff and Johnson 2003 [1980], 1999; Steen 2007, 2008; Kövecses 2002, 2005, 2015; Gibbs 2011) have gathered linguistic and experimental evidence that shows that metaphoric thought is central to reasoning and, as a result, pervasive and ubiquitous in human communication. This means that metaphor not only operates as a mere stylistic/ornamental and emotion-arousing tool in poetry and literature, but it underlies ordinary language and everyday (inter-)action. More specifically, metaphor has been shown to be a foundational, and thus, an essential instrument of the human mind to generate and structure thought and to get it across in communication (Gibbs 2011).

This finding has its roots in Experientialism, a philosophical theory of life that arose in the 1970s to take the place of the more traditional (and opposed) perspectives provided by Objectivism/Positivism and Subjectivism. The experientialist view contends that truth is relative to our conceptual system, which is grounded in our experience. In linguistic description, Experientialism was channelled through Cognitive Linguistics, a school that emerged in the early 1970s out of dissatisfaction with formal approaches to language (Evans and Green 2006, 2). Pioneered by Lakoff and Johnson (1980), Cognitive Linguistics holds that thought is largely based on imagery, i.e., metaphor, metonymy, and other figurative processes which go beyond the literal representation of reality (see also Lakoff 1987, 1993; Lakoff and Johnson 1999). These ideas lie at the origin of Conceptual Metaphor Theory, then expanded into what has been termed Conceptual Metaphor and Metonymy Theory (Dirven and Ruiz de Mendoza 2010, 38).

A conceptual metaphor consists in “understanding and experiencing one kind of thing in terms of another” (Lakoff and Johnson 1980, 5), where by “thing” is meant a domain of experience, which is structured in the form of an *idealised cognitive model*, i.e., a principled knowledge construct which captures aspects of our experience of the world. More specifically, conceptual metaphor involves *systematic mappings* (or sets of correspondences) between concrete and abstract conceptual domains (i.e., between the *source* and *target* domains, respectively) (Lakoff and Johnson 1980, 246). These metaphorical mappings, which structure our reasoning and our everyday language, “grow out of an *embodied* conceptualisation of reality”, that is, they arise from “the ways in which our species-specific *bodies* interact with the environment” (Evans and Green 2006, 27) [our emphasis]. In other words, everything that affects the nature of our experience, mind, and language cannot be studied in isolation from the configuration and conceptualisation of our bodies. This is what is known as *embodiment* (Lakoff and Johnson 1980, 55–56; 1999, 91).

For example, expressions from the field of Economics – such as “The value of shares in the New York Stock Exchange *rose* unexpectedly” and “Financial assets, among other stocks on the market, *plummeted* all of a sudden” – instantiate the metaphors MORE IS UP and LESS IS DOWN, respectively. As Lakoff and Johnson (1999, 47) explain, these conceptual metaphors arise because a subjective judgment of quantity is conceptualised in terms of the sensorimotor experience of verticality. This happens because of our experience with seeing levels either rise as larger and larger amounts of substances accumulate or drop as amounts gradually decrease.

Subsequently, other four major theories of metaphor, with a clear cognitive orientation, were developed.¹ Two of these theories were formulated within Cognitive Linguistics: (i) Blending Theory (BT) (e.g., Fauconnier and Turner 1996, 1998, 2002), which focuses on the conceptual structure that emerges from combining source and target inputs to metaphor; and (ii) Primary Metaphor Theory (Grady 1997a, 1997b, 1999, 2005; Grady and Johnson 2002), which argues that traditional metaphors can be reduced to more basic mappings that directly arise from primary experience (e.g., LOVE/LIFE/A CAREER IS A JOURNEY are enrichments of GOALS ARE DESTINATIONS). Outside Cognitive Linguistics, but also involving a cognitive orientation, is the Career of Metaphor Theory (e.g., Gentner and Rattermann 1991; Gentner and Bowdle 2001; Bowdle and Gentner 2005), which associates more conventional metaphors with cognitively more expensive analogical processes and less conventional ones with a more efficient categorisation of the target into a more abstract category cued by the source.

This chapter goes on to show in what ways and to what extent these cognitivist approaches to figurative thought and language have been useful to terminological metaphor description. At this point, it is worth mentioning that these theories lie at the heart of a growing body of studies that look into the nature of metaphor at all levels of linguistic description and explanation. For example, Lakoff (1987), Tyler and Evans (2003), and Evans (2004) take a lexical perspective, whereas Goldberg (1998) studies metaphor from a grammatical point of view, and Sperber and Wilson (1986) as well as Ruiz de Mendoza and Galera (2014) focus on the pragmatic aspects of figurative thought and language. Figurative linguistic instantiations can be addressed through the traditional qualitative, intuition-based methodology of metaphor examination or through a corpus-based/driven methodology (see Kövecses et al. 2019, for the advantages and disadvantages of the two methodological approaches).

2.2 Conceptual metaphor typologies

A programmatic account of conceptual metaphor classifies it into types. Lakoff and Johnson (1980, 3) were the first to suggest the basic types of verbal metaphor that can occur in speech and writing. Accordingly, metaphors are classified into three different groups: orientational, ontological, and structural metaphors. The first type of metaphor organises thought by providing a spatial orientation based on the configuration and functioning of our bodies (Lakoff and Johnson 1980, 14).

1. This chapter only mentions those theories that have been fruitfully applied in terminology studies. See Steen (2007) for a full account and detailed comparison of all theories of metaphor within Cognitive Linguistics.

Thus, orientational metaphors provide evidence that states of mind, feelings, and other abstract concepts can explicitly be related to physical orientations in order to explain the relation between our thoughts and movements verbally. As an example, Lakoff and Johnson (1980, 15) suggest the metaphors *HAPPY IS UP* and *SAD IS DOWN*, which are grounded in the empirical fact that we tend to slouch when feeling sad and stand and walk up straight when feeling happy. This rationale enables the use of metaphorical expressions in medicine, as in “Following pill administration, the concentration of free serotonin in the patient’s brain was *on the increase*, and thus, his mood/spirits *rose*”.

Ontological metaphors are a very helpful resource when explaining abstract concepts which are difficult to understand literally. In fact, they are naturally used to explain disembodied concepts such as emotions, states of mind and purely abstract ideas through objects, substances, containers, or people (Lakoff and Johnson 1980, 31). One common situation where we use ontological metaphors is the description of the human mind, which is understood in terms of a tangible object. For instance, we view states of mind as boxes into which we can put things and from which we can take things out by using metaphors such as “Due to sustained strain, the patient *fell into a depression*”.

Structural metaphors are the most complex ones because “understanding takes place in terms of entire domains of experience and not in terms of isolated concepts” (Lakoff and Johnson 1980, 117). This means that the whole or a substantial portion of the source domain is mapped onto the target domain. An illustrative example in the domain of civil engineering is *CONCRETE IS THE SKIN OF A CONSTRUCTED STRUCTURE*, which is pervasive in this specialised knowledge field (Boquera 2005, 198). This structural metaphor underlies a great many metaphorical terms such as “The concrete of our building is *weeping/bleeding* because it is running out of humidity” and “The cover of the old bridge is *alligatoring* due to excess vibration”. Spanish-language engineers usually make the following statement: “Cuando llueve abundantemente, el hormigón de poco espesor *se entumece*” [literally, “Whenever it rains heavily, thin concrete becomes *numb*”]. It can thus be concluded that this conceptual metaphor is recurrent in Western culture, where we are inclined to associate building materials with human skin and body parts. Therefore, the metaphor *CONCRETE IS THE SKIN OF A CONSTRUCTED STRUCTURE* is not used arbitrarily, but is the result of the impact of our cultural background. In this way, “we structure our experience – and it naturally emerges – from our activity in the world” (Lakoff and Johnson 1980, 119).

Unlike conventional (structural-conceptual) metaphors, image metaphors build upon one-shot comparisons, i.e., only one concept of the source domain maps onto the target domain, and thus, an image metaphor is conceptually simple (Lakoff 1993, 229; Lakoff and Turner 1989). Importantly, image metaphors

arise from sense-perceived resemblance (similarity), which involves comparison in terms of physical features such as shape, colour, and size. The image metaphor versus the structural-conceptual metaphor dichotomy is a correlation of the one-correspondence metaphor versus the many-correspondence metaphor typology suggested by Ruiz de Mendoza (1997). A one-correspondence metaphor results from a single cross-domain correspondence or mapping, and is thus not a part of a productive, many-correspondence metaphor.

The definition of image metaphor – and ultimately of resemblance metaphor – was later refined by Grady (1997a, 1997b, 1999). Based on the analysis of figurative expressions such as “Achilles is a lion”, Grady (1999, 89) distinguishes between behaviour-based metaphor in contrast to image metaphor. The former involves resemblance grounded in action/physical activity (in the previous example, Achilles’ courageous actions resemble the aggressive behaviour of lions) without no reference to physical appearance. However, the latter entails a comparison in shape, size or colour, thus based on static features. For example, the contour/shape of a river compared with that of a wool thread triggers the image metaphor *A RIVER IS A WOOL THREAD*.

Grady’s (1997a, 1997b) Primary Metaphor Theory differentiates between primary metaphors,² which are atomic or minimal structures that automatically and unconsciously arise from everyday experience by means of conflation, and compound metaphors, which are molecular (i.e., complex/composite) structures growing out of primary metaphors. Conflation (first suggested by Johnson 1997) refers to young children merging their subjective (non-sensorimotor) experiences and judgments, on the one hand, and sensorimotor experiences, on the other, so that they become undifferentiated and undistinguishable when they co-occur in everyday interaction (Lakoff and Johnson 1999, 46). Examples of these two types of metaphor are provided in Section 3.2.

Ruiz de Mendoza and Pérez (2011) take a multi-perspective approach to metaphor classification, based on the following parameters: (i) the ontological nature of the source domain, which elaborates on Lakoff and Johnson’s initial distinction between ontological, structural, and orientational, by expanding orientational metaphors into image-schematic metaphors, and then, by adding other categories like image metaphors (e.g., “A horse with a mane made of *rainbows*”, which uses specific images to provide a vivid picture of some aspects of the target), and situational metaphors (e.g., “He ran away with his *tail* between his *legs*”, which depicts a situation in the source); (ii) the nature of the mapping system (which incorporates the one-correspondence/many-correspondence distinction); (iii) the level of genericity

2. See Lakoff and Johnson (1999, 50–54) for a list of the most representative primary metaphors suggested by Grady (1997a).

of the target domain (which calls for a distinction between high and low-level mappings); (iv) the degree of complexity of the metaphoric operation (which expands Grady's distinction between primary and compound metaphors by contemplating the existence of metaphoric amalgams of various kinds, as in "She got the *idea across to me*", which combines into one IDEAS ARE MOVING OBJECTS and UNDERSTANDING IS GRASPING/SEEING).

Finally, there is the distinction between deliberate and non-deliberate metaphors, put forward by Steen (e.g., 2008, 2017) in his Deliberate Metaphor Theory. The former provides an innovative perspective on a topic. We find examples of deliberate metaphors in studies of molecular biology education and its terminology, as in the sentence "You could actually make sort of *Frankenstein* hybrid transcription factors with cloning techniques" (Beger 2016, 68). Using FRANKENSTEIN as a source domain for the process of separating and reassembling transcription factors is absolutely unconventional. In contrast, non-deliberate metaphors are not used as such by speakers, not even by specialists in diverse knowledge fields. For instance, the conceptual metaphor MARINE HABITATS ARE COMMUNITIES, which depicts marine life as a complex and competitive social structure, is not deliberately exploited as a metaphor by marine biologists. A clear example is seen in the following sentence, extracted from a marine biology research article: "In a sex-role reversed *population* of the *peacock blenny*, *Salaria pavo*, two alternative male types [...] mimic female-courtship behaviour and *nuptial* colouration in order to try to *sneak* fertilizations during spawning episodes" (Ureña 2011, 167).

On the basis of experimental evidence, however, Gibbs (2015) has argued that the notion of "non-deliberateness" in metaphor often lacks cognitive plausibility and should make explicit the pragmatic and conceptual signals which support their interpretation for adequate experimental testing (and consequent empirical support) to be possible.

3. Metaphor in Terminology and domain-specific language discourse

As previously mentioned, it was not until the late 1970s that figurative language began to gain momentum in scientific description and in the philosophy of science. Furthermore, in ordinary language research, cognitive linguists in the 1980s and 1990s, such as Lakoff (1987) and Turner and Fauconnier (1995), argued that metaphor, being primarily a matter of thought, also permeates the terminology of domain-specific and specialised language domains. Earlier on, metaphorical expressions were considered too vague and inaccurate to be used in specialised language discourse. This view is somewhat paradoxical, given the pivotal role which philosophers of science started to ascribe to metaphor and metonymy in

the development of scientific theories and in the description and explanation of complex concepts to laypeople and also to semi-specialised audiences (Boyd 1993; Hesse 1966; Kuhn 1979; Brown 2003 *inter alia*).

The lexical and conceptual nature of domain-specific metaphor as well as the interplay of the conceptual underpinnings and discursive (pragmatic and rhetorical) functions of metaphor have been explored in an extensive body of research all the way through to the 21st century either from a monolingual (mostly English) or inter-/cross-linguistic perspective. Much of this research now relies on the compilation, processing, and analysis of (large) computerised text corpus datasets by means of (semi-)automatic metaphor identification techniques. The standard theoretical framework for studies focusing on the lexical-conceptual dimension of specialised language metaphor is Conceptual Metaphor Theory. On this basis, metaphor has been investigated in diverse knowledge fields. For example, Brdar, Brdar-Szabó and Perak (2019) conduct a cross-linguistic analysis of transplant surgery metaphorical terms that reveals mismatches between the consistent pair Croatian–Hungarian, on the one hand, and English and French, on the other, in conceptualising domain-specific entities and processes (non-)metaphorically. This research highlights the suitability of constructing multi-lingual conceptual metaphor system repositories for unbiased, more integrative content description and representation.

By harnessing a civil engineering data corpus, Alexiev (2005) accounts for conceptual, linguistic and cultural differences and commonalities of metaphorical terms in English, Bulgarian and Spanish for translation purposes. By applying a set of highly productive strategies for the semi-automatic identification of marine biology terms in a corpus of specialised research articles in English and Spanish, Ureña and Faber (2011) reveal inter-linguistic similarities and differences in the conceptualisation of sea organisms through metaphor.

As for corpus-based, cognitive-oriented studies with a markedly discursive approach to domain-specific metaphor, many of them draw on Critical Discourse Analysis³ in conjunction with Conceptual Metaphor Theory and Blending Theory. For instance, in dealing with gender, Koller (2004) examines how *metaphor* can be employed to position people in *gender roles by analysing a political text corpus*. To address the topic of race based on conceptual blending premises, Hart (2008) looks at the metaphorical constructions of nation, immigration, and immigrants in election manifestos. In business and politics, Charteris-Black (2014) offers insightful guidelines to perform political speech analysis, where metaphoric thought and

3. This is an interdisciplinary approach to discourse that views language as a form of social practice.

language play a major role. Koller (2006) shows the suitability of using electronic text corpora to study metaphor in business media discourse.

Musolff (2016) discusses the widely debated political metaphors WAR, FAMILY, BODY and PERSON, identified in a bilingual English-German corpus collected from the press dealing with European politics. Sweetser, David and Stickles (2019) implement a methodology to detect metaphors from the target domain of TAXATION by analysing grammatical constructions extracted from corpora in English, Farsi, Russian, and Spanish. By analysing English-language tennis reports and online fora, Caballero (2012) sheds light on the *empathetic* and *evaluative* role of metaphor as a communicative strategy in sports discourse.

Once that terminological metaphor has been briefly presented, the following subsections describe how terminologists have studied metaphorical terms. In fact, they have been able (i) to make ground-breaking contributions to the understanding of conceptual metaphor by applying it to terminology; and in some cases (ii) to refine the metaphor types initially proposed by general language metaphor researchers.

3.1 Challenging Boyd's typology: Novel and deliberate metaphors in scientific circles

As previously mentioned, philosophers of science aptly claim that "metaphor is important in theorising and can be a part of the theory" (Hoffman 1980, 416). Boyd (1993) makes an important contribution to philosophy of science by distinguishing between *theory-constitutive metaphors*, *heuristic metaphors* and *exegetical/pedagogical metaphors*. The first ones are intrinsic and foundational to certain scientific theories and formulations, and thus, they cannot be paraphrased in literal terms. A stock example is the metaphor THE HUMAN BRAIN IS A COMPUTER, which gives rise to statements such as "Thought is a type of *information processing*" and "Learning is an adaptive response of an *auto-controlled machine*" (Boyd 1993, 486).

Though not pre-theoretical, heuristic metaphors help develop theories and support scientific knowledge by solving or presenting new approaches to problems specific to each specialised discourse. They are used when it is difficult to express theoretical premises literally, or even when direct observation is not possible. The expression "information highway", used to refer to the Internet, is an instance of heuristic metaphor. Both theory-constitutive and heuristic metaphors are linguistically realised as metaphorical *terms*, namely, well-entrenched expressions with specialised meaning in the lexicon of scientific discourse.

Finally, exegetical metaphors do not transmit new scientific thought. They only play a role in "the teaching and explication of theories of science and technology

which already admit of entirely adequate non-metaphorical (or at any rate less metaphorical) formulations” (Boyd 1993, 359). Examples of pedagogical metaphors are “worm-holes” in general relativity, the description of the spatial localization of bound electrons in terms of an “electron cloud”, or the description of atoms as “miniature solar systems” (*ibid.*).

Because metaphor had initially been regarded as inexistent in scientific circles, systematic cognitive-linguistic research on metaphor in subject-oriented and specialised knowledge domains only began in the 1990s. This fact explains why embryonic studies of metaphor in diverse fields of expertise clung to Boyd’s (1993) *watertight* classification of metaphors in science and technology. Research in this vein was conducted in the fields of Computer Science (Meyer, Zaluski and Mackintosh 1997), economics (White-Hayes 1996), medicine (Tercedor 1999), publicity (Cortés de los Ríos 2001), and science teaching (Bradford and Dana 1996), to name but a few.

Challenging Boyd’s (1993) monolithic metaphor typology in specialised communication, Knudsen (2003, 2005) notes that the dividing line between theory-structuring and pedagogical metaphors is fuzzier than initially assumed. For example, Knudsen (2003, 1261) shows that the GENES ARE A LINGUISTIC CODE metaphor (e.g., “Once a gene has been *transcribed* into *messenger* RNA, those unwanted stretches of *transcript* have to be *deleted*”) started out as a pedagogical (i.e., science popularising) metaphor, which was transformed into a theory-constructive one, and now is generally used again in non-expert communicative settings as well.

In the same direction, Ureña (2016) draws on premises from the Career of Metaphor Theory (Bowdle and Gentner 2005) to analyse empirical textual data from academic marine biology journal articles. Ureña (2016) demonstrates that novel and spontaneous metaphors are sometimes used by biologists to address their peers in formal expert-expert communication with a view to: (i) explaining and describing findings in the field in a highly creative – and at the same time – effective manner; and (ii) humorously bringing the readership’s attention to the claims made in their articles with playful analogies (rhetorical and discursive purpose). In this study, the array of novel metaphors is shown to be systematically exploited in order to conceptually structure and give thematic coherence to a whole specialised discourse event dealing with school fish behaviour. These creative metaphors, some of which are grounded in metonymy, may be either non-terminological lexical items and expressions or terminological units borrowed from other specialised knowledge fields.

An example of non-terminological metaphors identified in Ureña (2016) is “*club* members do not *lift a fin(ger)*”. This metaphorical sentence is used by the author of the article to explain that whenever a fish school has a “free rider” (individual fish which advances under water on the swimming efforts of its schooling

partners or “*fellow club members*”), this fish will be automatically excluded by guard fish with the implicit consent of the rest of school members. The author of the article uses the informal, ordinary language idiom “not to lift a finger”, which means “not making any effort at all to do something, especially to help somebody”, to construct an adapted version of this idiom that exploits the pun “fin(ger)” [fin (fish body part) + finger (human organ)]. The interesting aspect of this rhetorical figurative device is that it is used to describe school fish behaviour and dynamics (domain-specific information) in a specialised knowledge domain context.

In the field of Ecology, Thayer (1979) suggests the novel terminological metaphor “biological bulldozer”, a phrase borrowed from civil engineering terminology, to refer to those bioturbators that modify and shape marine ecosystem surfaces (such as the seabed) by burrowing, ingestion, and defecation of sediment grains. This once creative metaphor became a well-entrenched term when – partially inspired by Thayer – Jones, Lawton, Shachak (1994) systematically developed the Ecosystem Engineering theory. On this basis, they proposed the terminological unit, *ecosystem engineers*, and began to use the metaphorical noun “bulldozer” in verbalised forms, as in the sentence, “The European *periwinkle* (*Littorina littorea*) bulldozes sediments from hard substrates” (Jones, Lawton and Shachak 1994, 376). As they put it, this terminology was necessary since “the role of organisms in the creation and maintenance of habitats had previously neither been defined nor systematically identified and studied” (Jones, Lawton and Shachak 1994, 373).

Novel metaphoric expressions *deliberately* exploited in specialised scientific settings are thus closely related to the notion of *deliberate metaphor* suggested by Steen (2008, 2017), which was defined in Section 2.2. On this basis, Ferella (2020) analyses deliberate metaphor usage in three Hippocratic treatises and in Empedocles’ two famous similes. Ferella examines direct metaphors – as she also names them – through their textual markers with a view to testing the illustrative as well as heuristic communicative functions of these metaphors in ancient Greek anatomical and physiological dissertations. She concludes that these are deliberate metaphors because the authors who exploit them claim to use them as metaphors, leaving their addressees no option but to recognise the source domain as such (Ferella 2020, 129). For example, a Hippocratic author establishes the novel deliberate metaphors, FEMALE WOMB IS AN OVEN, FOETUS GROWTH PROCESS IS BAKING, and MEMBRANES ENCOMPASSING THE SEED (FOETUS) ARE THE CRUST OF BREAD (there is scientific evidence that the foetus forms a membrane around itself when it inflates with air; hence the *membranes-crust of bread* comparison). This network of spontaneous, deliberate metaphors comes into being because the maternal uterus is thought to be hot (Ferella 2020, 136).

3.2 Compound and primary metaphors and image-schematic metaphors

As mentioned in Section 2.2, research on metaphors in subject-oriented vocabulary also concerns compound, and thus, primary metaphors. An example of compound metaphor in politics and business is *COMPLEX ABSTRACT SYSTEMS ARE PLANTS*, which arises from the primary metaphor *ABSTRACT COMPLEX SYSTEMS ARE PHYSICAL OBJECTS*. Accordingly, we can build sentences such as “Please turn to the local *branch* of the organisation” and “Government and educational bureaucracies should be *pruned*” (Kövecses 2005, 209, 210). In the same vein, Taylor and Van Every’s (2000) formulation of scientific management, which involves a mixture of mechanical engineering and social physics, conceives social behaviour in organisations as following deterministic, natural patterns; that is, workers need to be selected on the basis of their adequacy to a particular pre-specified task and can be characterized in units of energy – hence the compound metaphor *ORGANISATION WORKERS ARE FOOT-POUNDS OF ENERGY* (Cornelissen and Kafouros 2008, 966). Taylor is thus using the conventional primary metaphor *ORGANISATION IS A PHYSICAL STRUCTURE* (a machine, in this case) to conceptualise an organisation in terms of the concrete structure of an object, with physical parts, which in turn represent the different elements of the organisation.

From a monolingual perspective, Yu and Tay (2020) draw on Corpus Linguistics as well as on discourse and statistical analysis to eventually find a strong correlation between image-schematic metaphors and the themes of anger, anxiety, and depression in an English-language specialised corpus of therapeutic transcripts. Image schemas such as *FORCE* (e.g., “It is hard to *pin* anger *down*”) and *VERTICAL ORIENTATION* (“Anxiety *builds up*”) are shown to be among the commonest types of image-schematic metaphors used by patients to describe the therapeutic themes of anger, anxiety, and depression. Importantly, this research has implications for how image-schematic metaphors can be used to facilitate the descriptions of the above-mentioned topics in therapeutic conversations (Yu and Tay 2020, 253).

3.3 Refining the theory of resemblance and non-resemblance metaphors in specialised language

Specialised language metaphor theory has substantially benefited from studies of conceptual metaphor either from a lexical (terminological meaning) or a discursive (rhetoric and genre) perspective. Alexiev (2005, 40) explains that “conceptual aspects of terminological metaphor involve the mechanism of metaphorization, key conceptual themes underlying the metaphorical terms in a given terminology, motivation for metaphorical transfer and distinction between conventional (conceptual) and image (one-shot) metaphors”. Important refinements concerning the

two latter points are made by Ureña and Faber (2010) in their lexical semantic study of marine biology terms and Caballero (2006), who explores the creative power of metaphors in architecture from a conceptual and rhetorical perspective.

In regard to image metaphors in Terminology, Ureña and Faber (2010) remodel the traditional watertight sub-category division between image metaphors and behaviour-based metaphors (Grady 1999) by carefully considering the motivations for metaphorical transfer underlying terminological designations of sea organisms. Based on a cognitivist approach to such terms, Ureña and Faber (2010) show that there is not only a clear-cut distinction between metaphors based on motionless images (involving the static physical attributes of entities, such as colour, shape or size, as in “sea lettuce” and “whale shark”), but also between metaphors grounded in dynamic structure (i.e., function and motion-induced behaviour, as in “hermit crab”). They find that there are also metaphors that remain in a transitional area between behaviour and purely physical features, thus sharing aspects from both dimensions.

For instance, the comparison between the source and target in the metaphorical term *anglerfish* arises not only from behaviour (dynamic structure referring to the way the fish lures its prey and pulls at its fleshy, fishing line-like protuberance on its head), but also from physical resemblance of this protuberance to the fishing line (static features). Another example is the term “boxer crab” (species *Lybia Tesellatta*) (Ureña and Faber 2010, 9). This crab resembles a boxer inasmuch as it uses anemones attached to its claws (comparison in shape and function) to defend itself against octopuses in the same way as a boxer uses his fists to punch his opponent (comparison in behaviour).

Ureña (2011) as well as Ureña and Faber (2010) use the umbrella term *resemblance metaphors* to encompass all metaphor types that involve a comparison in shape, and/or size, and/or colour and/or behaviour. They use the term *non-resemblance metaphors* for those figurative compositions that put abstract patterns into play, along the lines of Lakoff and Johnson’s (1980) structural metaphors. Based on a marine biology corpus analysis, Ureña (2011) formulates standard multiple-correspondence non-resemblance metaphors. A case in point is SEA ACTIVITIES ARE ECONOMIC AND BUSINESS AFFAIRS, which could also be formulated as SEA ACTIVITY IS BUSINESS for short. This metaphor underlies many metaphorical uses, such as “energetic cost (of biochemical processes)”, “expensive (copulation) investment”, “(high nutrient) pay-off”, “(benthic food) budget”, “(the epifauna) capitalises on (local diversity)”, “net photosynthesis”, and “consumer organism”.

In comparing the two metaphor categories, however, Ureña and Faber (2010) also conclude that the dichotomy *resemblance metaphor vs. conceptual metaphor* is oversimplistic since being conceptual in nature, resemblance metaphors can also activate multiple cross-domain mappings of rich knowledge and productive

inferential structure, features initially attributed to structural metaphors (Lakoff and Turner 1989, 91). The reason for this is the nature of mental imagery, which is an integral part of every stage of metaphor formation (Ureña and Faber 2010, 143).

For example, marine biology resemblance metaphor terms such as “*surgeonfish*”, “*anglerfish*”, “*pilotfish*”, “*Spanish dancer*” (a nudibranch/sea slug species) and “*inn-keeper worm*” can be abstracted into the conceptual metaphor SEA ORGANISMS ARE JOB HOLDERS. This conceptual metaphor further licenses linguistic expressions used by experts to designate marine biology concepts, which are closely related to the abovementioned sea animals. As Ureña and Faber (2010, 137) argue, these linguistic expressions may emerge either from sheer physical resemblance (appearance) or from comparison in behaviour – kinetic structure that can be considered to be working as a more abstract *tertium comparationis*. Accordingly, surgeonfish have “scalpels” (term referring to sharp erectile spines near the base of these animals’ tail), which are used for scaring potential predators away or tearing their skin if attacked; anglerfish use their fleshy “baits” as a lure to catch prey; harvestfish “harvest” food for survival; pilotfish “cruise” long distances tagged alongside sharks in a mutualistic relationship; Spanish dancer nudibranchs undulate their “ribbons” (term used by biologists to designate sea-slug mantle spirals) to swim forward and backwards across the sea water; and the burrow of inn-keeper worms is occupied by several “commensals”.

The claim that resemblance and non-resemblance metaphors cannot always be classified as separate categories also finds support in general-language metaphor research (e.g., Deignan 2007) and as well as in architecture (Caballero 2006). In her corpus-based study of the conceptual and rhetorical roles of metaphorical, fictive-motion verbs⁴ used by architects to assess design solutions in building reviews, Caballero (2006) suggests a body of conceptual metaphors which reveal their highly imagistic nature while also being conceptually complex. For example, Caballero (2006, 122) finds MALLEABILITY to be a recurrent schema licensing metaphoric thought in architectural structure description. This schema conceptualises and characterises buildings and parts of them as pliable (“fold”, “unfold”), soluble (“melt”) or flexible solids (“stretch”, “splay”, “spread”, “flex”, “embrace”). An illustrative corpus example is the sentence “Large departments *embrace* two atria and each atrium can be enlarged to form *nodes*” (Caballero 2006, 153).

These metaphors are imagistic and are thus grounded in physical resemblance because the form that they evoke matches the actual shape of the entities described.

4. These are motion verbs used to describe spatial configurations which do not involve actual motion or change. Talmy (1996, 211), the first scholar to address fictive motion, defines it as “linguistic instances that depict motion with no physical occurrence”. More specifically, fictive motion is simulated motion arising from our visual scanning mechanisms (Talmy 2000; Matlock 2004).

This comparison is ultimately based on how the entities are visually scanned by virtue of what Caballero (2006, 3) calls *the expert's visual thinking*. At the same time, these metaphors can all be regarded as instantiations of the more general metaphor FORM IS MOTION, formulated by Lakoff and Turner (1989, 142–144) to account for a wide array of metaphorical expressions that convey fictive motion in everyday communication (e.g., “The path *stretches along* the shore of the lake” and “The fence *dips and rises* in parallel with the terrain”).

It can thus be concluded that underlying metaphoric thought captured by terms of specialised language (e.g., marine biology and architecture) there is a harmonious interaction between the conceptual and the purely perceptual attributes of the entities under comparison.

3.4 Socio-cultural and contrastive aspects of terminological metaphor

There is broad agreement among cognitive linguists that metaphorical thought is the result of an inescapable combination of “nature and nurture” (Sinha 1999). This assumption was firstly acknowledged in ordinary language. Because they were shown to essentially determine the shared biology and bodily experiences that give rise to metaphorical thought, socio-cultural factors underlying figurative reasoning eventually received richly deserved attention as from the beginning of the 21st century (cf. Kövecses 2002, 2005, 2015; Dirven et al. 2008; Rohrer 2006; Sinha 1999; Zinken, Hellsten, and Nerlich 2003, *inter alia*).

For example, the non-universal (culturally restricted) metaphorical conceptualisation of emotions such as ANGER, LOVE, and HAPPINESS has been discussed from a cross-cultural perspective (see Kövecses 2005 for a contrastive study of English, Spanish, Bulgarian and Chinese lexical units; Yu 2008 for an English-Chinese comparative analysis of metaphor, body and culture). Socio-cultural factors constraining metaphor-induced conceptualisations of reality give rise to cross-linguistic differences in everyday discourse. In fact, “there can be differences in the range of conceptual metaphors (or, more precisely, the range of source domains) that languages and cultures have available for the conceptualisation of particular target domains” (Kövecses 2005, 70). Most importantly, scholars have eventually concurred that a comprehensive, well-balanced view of metaphor calls for a *joint* intervention and analysis of the “situational, discourse, bodily, and conceptual-cognitive contexts in which cognition is not only embodied but also grounded in multiple ways” (Kövecses 2015, xii).

As with ordinary language research, there are cognitively oriented studies that account for the socio-cultural factors impinging on specialised concept formation through metaphoric thinking. Temmerman's (e.g., 2000, 2002, 2008) Sociocognitive Theory of Terminology is the first one to emphasize conceptual organisation from

a cognitive-linguistic perspective with a view to describing the intrinsically polysemic and variable nature of the terminology of specialised knowledge domains. In fact, Temmerman argues for the interdisciplinarity of terms, thus rejecting the assumption in the General Theory of Terminology (Wüster 1979) that a term should only belong to one knowledge field.

Temmerman's proposal is especially inspirational to cognitivist studies of terminological metaphor because it explores the influence of sociocultural situatedness on metaphorisation, resulting in extensive terminological metaphor usage in the specialised discourses of genetics and biotechnology. Much along the lines of Boyd (1993) in the philosophy of science, Temmerman claims that metaphorical thinking is a salient type of thought in science which is essential for scientific advance. As she highlights, "life specialists understand by means of metaphorical domains and [...] these metaphorical domains have left their traces in language as lexeme metaphors (Temmerman 2000, 180). Temmerman (2008) illustrates the flexible nature of metaphor and the role of category prototypicality and extension as a functional tool to fulfil scientists' needs for linguistic economy and scientific term creation. She supports this claim through the historical, diachronic, semantic and discourse analysis of the polysemous term "splicing", which started out as a non-terminological word (e.g., "rope splicing", "film splicing") and has now become part of the specialised terminology of genetics (e.g., "gene splicing", "mRNA splicing").

In relation to the inter- and cross-linguistic studies of terminology, Alexiev (2005, 38) claims that "the choice of a target language conceptualisation strategy and the corresponding surface realisation depend both on cognitive and language- and culture-specific factors". In his cross-linguistic study of Bulgarian, English and Spanish terms in the fields of mining, geology and civil engineering, Alexiev (2005, 108–115) establishes a set of cultural-experiential parameters that determine the choice of the designation and general concept to be exploited in the terminological metaphorisation process. On the basis of data from marine biology, Ureña and Tercedor (2011) refine Alexiev's classification to offer a typology that sheds light on the relationship between the physical and the sociocultural underpinnings of terminological resemblance metaphor. Specifically, Ureña and Tercedor's (2011, 216) research builds upon two major claims: (i) sensorimotor perception and sociocultural factors merge into a physical-social experience that shapes scientific knowledge through metaphor; and (ii) the sociocognitive patterns involved in terminological metaphor formation give rise to inter-linguistic variation and commonalities.

After comparing English-Spanish term pairs, Ureña and Tercedor (2011) propose a typology of terminological metaphors based on their level of sociocognitive situatedness. The vast majority of these metaphors are of the resemblance type since they involve the physical comparison of the source and target (similar

superficial appearance). Specifically, there are (i) culture-specific metaphors, which are further divided into three subgroups (see Ureña and Tercedor 2011, 223); (ii) culture-typical metaphors; (iii) metaphors arising from unconstrained angles of referent perception; and (iv) metaphors showing varying degrees of specificity. Hard-core instances of culture-specific metaphor in marine biology exhibit culturally unique source domains in one of the two languages under comparison.

A case in point is the Spanish term “cangrejo *moro*” [literally “*Moorish* crab”], where the dark skin colour of a *moro* (a term exclusively used by Spaniards in the VIII century to designate Muslims who had invaded and settled in Spain at that time) is compared with the darkened exoskeleton of this crab species. The English-language equivalent is “sally *lightfoot* crab”, a metonymic and metaphorical term whose figurative nature is not culturally motivated. The metaphor is found in “foot”, a word literally denoting the lowest part of a human’s leg, which is used to refer to a crab’s walking leg.⁵ This metaphor is thus based on a part-for-whole metonymy, where the part (FOOT) stands for the whole (LEG). The noun “sally” and the adjective “light” refer to the physical ability of this crab species to rush out of rock cracks and holes in order to very quickly escape predators or chase prey.

Culture-typical metaphors feature concepts that are characteristic of a community of speakers. For example, the term “*cookie-cutter* shark” arises from the comparison between the cookie-shaped plugs of flesh extracted by the shark from its prey and the actual shape rendered by a cookie-cutter on cookie dough. As is widely known, cookie-cutters are popular baking utensils in the English world (typically in American pastry-making). In contrast, the Spanish-language equivalent used by marine biologists is “*tiburón cigarro*” [“*cigarette* shark”] because of its thin, elongated shape and the dark, collar-like mark around the base of its head, which resembles the tip of a cigarette (Ureña and Tercedor 2011, 226–227). Cigarettes can be found all over the world, so they cannot be labelled as culture-typical, much less as culture-specific.⁶

Demaeker (2017) analyses the linguistic and conceptual nature of corpus-retrieved wine-tasting metaphorical terms from an English-French inter-linguistic perspective. She does so with a view to building translation-oriented dictionaries and providing specialised language translators and terminologists with useful tools to carry out their work. Demaeker (2017) demonstrates that wine-tasting conceptual metaphors can be analysed and characterised by means of

5. A crab has three major limb types: walking legs, swimming legs and chelipeds, which include claws and pincers.

6. Due to space restrictions, only the first two metaphor types in Ureña and Tercedor’s (2011) classification are briefly discussed in this paper. See these authors’ article for a full account of all metaphor categories in their typology.

conceptual integration – that is, through blending mechanisms. In comparing linguistic expressions in the languages under analysis, in this case, English and French, the two languages may rely on the same conceptual metaphor, but sometimes the metaphoric equivalents do not perfectly correspond. By way of example, Figure 1 shows the integrated schema of the English and French lexical items “enclosing” and “s’appuyant sur”, which can be subsumed under the metaphor THE TASTE OF WINE IS A CONSTRUCTION.

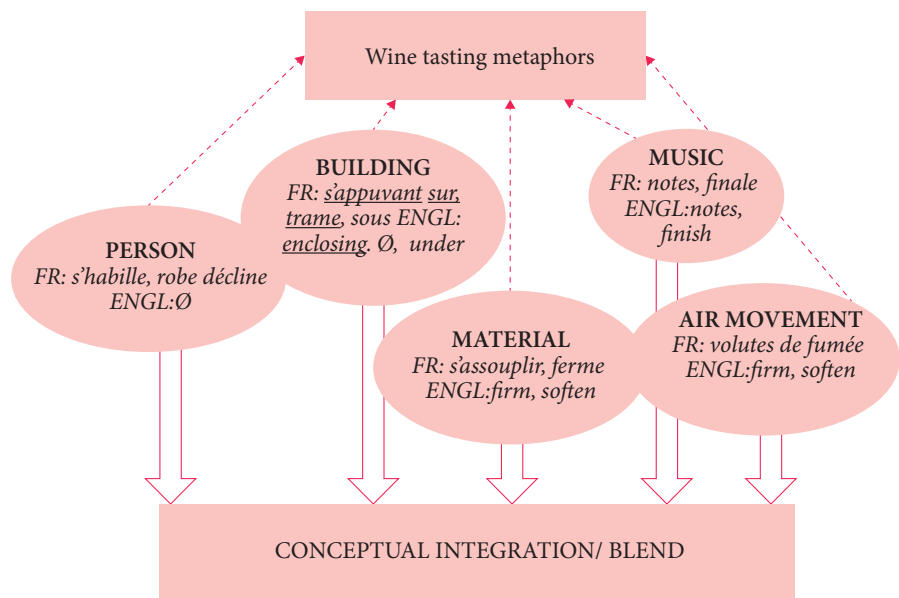


Figure 1. Linguistic expressions of the source domains in a conceptual blending structure, which describes the wine-tasting metaphors “s’appuyant sur” and “enclosing” (Demaecker 2017, 127)

Differences in concept construal through metaphor have also been documented in specialised language discourse from an *intralinguistic* perspective. Ureña and Faber (2014) compare resemblance metaphor terms designating sea organisms that were semi-automatically retrieved from a subcorpus of Spanish-language marine biology research articles. These articles were written by marine biologists in a wide range of Spanish language varieties, including those occurring in Latin-American countries and in Spain (European/Peninsular Spanish). The analysis of this authentic data revealed interesting cases of language-internal terminological diversity not only across the two continents but also within one same country. An example of inter-continental variation is the pair “delfín nariz de botella” [bottlenose dolphin] – “delfín mular” [mule dolphin]. Both terminological units refer to the

dolphin species *Tursiops truncatus*. However, the metaphorical term “delfin nariz de botella” is chiefly used by Latin American biologists. The metaphor is based on resemblance in shape between this animal’s snout, which is particularly short and stubby, and a bottle. In contrast, the corpus data demonstrates that Spanish biologists prefer to use the metaphorical term “delfin mular” because of this animal’s robust appearance as well as its energetic and hardy nature, which is reminiscent of that of a mule (Ureña and Faber 2014, 205).

Geographical fragmentation is thus shown to reflect differing sociocultural models even in one single language. These different models have a direct bearing on the conceptualisation and denomination of entities, attributes, processes, and events in specialised knowledge fields, as is the case of marine biology. Hence the emergence of intra-linguistic denominative variation in science.

4. Conclusions

This chapter has offered an overview of the emergence and development of conceptual metaphor from the perspective of Cognitive Linguistics both in general language and in the terminology of specialised knowledge domains. By disabling wrong assumptions from past theories of language, Cognitive Linguistics provides empirical evidence that metaphorical thought is central to reasoning and, as a result, pervasive and ubiquitous in all spheres of human communication. Ever since the advent of this theoretical framework in the late 1970s, conceptual metaphor has been analysed and described at all levels of verbal description.

Terminologists have studied metaphorical terms and made ground-breaking contributions to the understanding of conceptual metaphor by applying it to specialised language. Importantly, terminological metaphor researchers have refined some of the metaphor types first suggested by general language metaphor scholars. For instance, besides the two classical kinds of resemblance metaphor, behaviour-based and image metaphors, Ureña and Faber (2010) study marine biology terms to show that there are resemblance metaphors remaining in a translational zone between the two mentioned above, thus sharing aspects of both. In architecture, Caballero (2006) provides corpus-based evidence that image metaphors can exhibit conceptual structure, just as much as conventional conceptual metaphors.

Boyd’s (1993) classification of metaphors in science shows that metaphorical terms are essential to create and further scientific and technological knowledge, whereas exegetical (i.e., novel) metaphors are exclusively used to explain specialised concepts to laymen. However, in identifying the different roles and functions of figurative thought in the domain-specific discourses of genetics and marine biology, Knudsen (2005) and Ureña (2016) refute Boyd’s (1993) monolithic division

between purely terminological and novel/deliberate metaphors. These scholars demonstrate that totally spontaneous and unconventional metaphors are frequent in expert-to-expert communication, not only to humorously describe specialised concepts, but also to conceptually and linguistically structure full scientific units of discourse and advance science and technology. Therefore, challenging previous theory, terminologists have shown that subject-oriented, expert-to-expert communicative contexts are fraught with novel metaphors, which are exploited by specialists for conceptual, textual, and rhetorical (attention-grabbing, imagination-stimulating, and evaluative) purposes.

Terminological metaphor research is also of great use to Cognitive Linguistics because it shows that the embodied (conceptual, cognitive, and bodily) underpinnings of metaphor are indissolubly linked to underlying sociocultural factors. These factors dramatically constrain the conceptualisation and linguistic realisation of entities, attributes, processes, and events in specialised knowledge domains. The merging of physical and social experiences involved in terminological metaphor formation gives rise to inter- and intralingual variation. For this reason, there are studies that provide typologies of terminological metaphors based on their level of sociocognitive situatedness.

Notwithstanding, terminological metaphor research still has much left to discover. One future challenge is to accommodate current burgeoning conceptual metaphor research from a Cognitive Semiotics perspective into terminology studies. This line of investigation outside Cognitive Linguistics, known as *Motivation and Sedimentation Model* (MSM), builds on ideas from classic thinkers, from Aristotle to Merleau-Ponty. MSM argues that metaphors should continue to be treated primarily as signs and sign processes, both from a mono- and multilingual perspective, and that the key point is to distinguish between motivation, convention and situated use, all of which are necessary in order to make sense of the complex dynamics of metaphor (Zlatev, Jacobsson and Paju, 2021). MSM has as yet only been applied to metaphor in ordinary language (Devolder and Zlatev 2020; Zlatev, Blomberg and Magnusson 2012; Zlatev, Jacobsson and Paju, 2021). There is thus an extensive testing ground for its implementation in the terminology of domain-specific and subject-oriented discourses. Studies in this direction show great promise and will doubtlessly yield valuable insights into the nature of terminological metaphor.

PART 7

Variation and equivalence

Causes of terminological variation

Judit Freixa

Campus del Poblenou

The aim of studying terminological variation is to understand the reasons for the existence of several denominations for the same concept in specialised texts. Based on Freixa (2006), this article resumes the study of the causes of terminological variation and updates the proposal with a literature review of the research on this topic in recent years. We first situate terminological variation as a natural and necessary feature of terminological units. We then give an overview of the advances in studying and representing variation and its causes. This leads to the main part of the article, which reviews the different causes of variation and provides new nuances and examples in light of the literature reviewed.

Keywords: variation, Terminology, causes of variation, terminological variation, denominative variation, cognitive variation, diachronic variation, diatopic variation, specialised discourse

1. Introduction

Like all other crises, whether political, socioeconomic, medical or environmental, the global pandemic caused by the SARS-CoV-2 virus has led to the emergence of new terms that speakers use. While some of them already existed in the strictly epidemiological sphere, others have since arisen to denominate new concepts, and many of them have been disseminated in non-specialised communications. All countries and all languages have had to face up to these new denominative needs in order to find ways of ensuring that terms related to the pandemic can be adapted to the diversity of communication scenarios in today's society. Within this context, terminological variation has played an important role, as we can see, for example, from record 1353023 in the IATE (2020) database, where the **equipment designed to protect people from catching the virus** is denominated by seven different terms in English: *personal protective equipment*, *PPE*, *personal protection equipment*, *personal safety equipment*, *personal protective device*, *individual protective equipment*

and *protective equipment*. Neither these synonymous terms nor any other example of denominative variation in any specialised knowledge area are randomly distributed, and trying to understand what social, cognitive and discursive factors they respond to is what drives the study of the causes of variation.

This chapter resumes the study of the causes of terminological variation by taking the proposal in Freixa (2006) as the starting point and then revising it mainly in light of the contributions made by the considerable work on terminological variation in the past two decades. Going beyond descriptive aims, and with ever greater rigour and knowledge, these works attempt to explain the patterns governing denominative alternation by comparing languages or by diachronic, sociolinguistic, or cognitive linguistic studies.

We firstly give a brief presentation of terminological variation as a natural and necessary feature of terminological units, and then provide an overview of the advances in the study and representation of variation and its causes. The main sections of the chapter focus on the causes of variation, and provide new nuances and examples in light of the literature reviewed.

2. Terminological variation: Natural and necessary

In General Linguistics, the lexical relationship between words having the same meaning is known as *synonymy*. Freixa (2006) proposes that this phenomenon in Terminology should be called *denominative variation* because synonymy is usually understood as being between different lexemes, whereas the term *denominative variation* covers any other type of variation in denominations¹ – from graphical, orthographical, morphological and morphosyntactic variations and reductions to lexical variations – so long as they can be considered *denominations*, in other words, not paraphrases or occasional occurrences but lexicalised forms with stability and consensus among the users of units in a specialised domain (Freixa 2006, 51).

The basis on which Kleiber (1984) defends the distinction between denomination and designation is referential fixation, which is established in denominations but not in designations:

1. Thus, taking the previous example of *personal protective equipment*, synonyms would be those names with a lexical change (such as *personal protective device*, *personal safety equipment*), but we also consider the acronym *PPE*, reduced forms (such as *protective device*) and names with morphological variation (*personal protection device*) to be denominative variation of *personal protective equipment*.

Pour que l'on puisse dire d'une relation signe < – > chose qu'il s'agit d'une relation de dénomination, il faut au préalable qu'un lien référentiel particulier ait été instauré entre l'objet x, quel qu'il soit, et le signe X. Nous parlerons pour cette fixation référentielle, qu'elle soit le résultat d'un acte de dénomination effectif ou celui d'une habitude associative, d'acte de dénomination, et postulerons donc qu'il n'y a relation de dénomination entre x et X que s'il y a eu un acte de dénomination préalable. (Kleiber 1984, 79)²

In contrast, Wüster argued that denomination was a type of designation, alongside symbols and other types of signs that were designations but not denominations (Wüster [1959–1960] 2001, 42).

In any event, *terminological variation* is also commonly used to refer to *denominative variation*, as in this chapter, because the variation of denominations given to a specialised concept is, in fact, the type of terminological variation that is most often addressed in the literature. Strictly speaking, denominative variation is a type of terminological variation because terms are affected by variation in different ways, and not just formally. In addition, conceptual variation is a type of terminological variation. Taking this even further, Condamines (2010) argues that periphrastic and designative variations for a concept are also examples of terminological variation, alongside variations in the use of terms correlated with textual genres and in the distribution of terms in texts.

The truth is that variation in Terminology is now fully accepted, thanks to the debate initiated on this topic by pioneering contributions such as Kocourek (1968); De Bessé (1974); Duchacek (1979); Natanson (1981) and Lethuiller (1989). They laid the groundwork for research on its causes by Daille et al. (1996), Mortureux (1997), and Bowker (1998), who, more than 20 years ago, came to the following conclusion:

[...] in other words, an expert does not choose a term based on principles of univocity or grammatical correctness, but rather based on factors such as the purpose or the focus of the communication situation at hand. An expert may consciously choose to use one term to emphasize a particular dimension at one time, and another term to focus on another dimension at another time, even though both terms are referring to the same concept. Moreover, we submit that this process is both valid and necessary in specialist communication, where there is an emphasis on precision. (Bowker 1998, 494)

2. To be able to say that a sign < – > thing relationship is one of denomination, a particular referential link must firstly be established between object x, whatever it might be, and sign X. For such referential fixation, be it the outcome of an act of effective denomination or that of an associative habit, we will speak of an *act of denomination* and will therefore postulate that there is a relationship of denomination between x and X only if there has been a prior act of denomination (our translation).

Since then, every theoretical proposal has tried to account for the real operation of terms by taking into consideration the diversity of scenarios in which terms occur. These proposals are often based on recent advances, including – but not limited to – Cognitive Linguistics, Sociolinguistics and Neurolinguistics, being Socioterminology (Gaudin, 2003) the first one to make this a fundamental issue to account for. The Communicative Theory of Terminology (CTT) thus contains the principle of variation among its fundamental postulates:

Todo proceso de comunicación comporta inherentemente la variación, explicitada en formas alternativas de denominación del mismo concepto (sinonimia) o en apertura significativa de una misma forma (polisemia). Este principio es universal para las unidades terminológicas, si bien admite diferentes grados según las condiciones de la situación comunicativa. El grado máximo de variación de la terminología lo cumplirían los términos de las áreas más banalizadas del saber y los que se utilizarían en el discurso de registro comunicativo de divulgación de la ciencia y de la técnica; el grado mínimo de la variación sería propio de la terminología normalizada por comisiones de expertos; el grado intermedio, la terminología usada en la comunicación natural entre especialistas. (Cabré 1999a, 85)³

In the new millennium, terminological variation has been deemed natural and necessary, and the aims of research have become more diverse in order to make progress on different fronts. Such research includes empirical work aimed at getting a better understanding, not only of the rules governing the operation of terminological variation, but also of its theoretical underpinnings. Freixa and Fernández-Silva (2017, 176) defend the idea of the unsaturability of concepts as one of the foundations of terminological variation: “...we defend the hypothesis that concepts are unsaturable with the understanding that concepts are constantly under construction, and that terms are not limited to the preferred form used to refer to a concept”. The study of the causes of variation is unquestionably the focal point of such research, since it is a matter of finding out not only why, how and when terms vary, but also who is responsible for such variation.

3. Every communication process inherently leads to variation, which is explicit in alternative forms of denomination of the same concept (synonymy) or in the openness of meaning of the same form (polysemy). This principle is universal for terminological units, although it admits different degrees depending on the circumstances of each particular communicative situation. Terms pertaining to the most ordinary areas of knowledge and those used in discourse having a communicative register particular to popular science and technology would exhibit the highest degree of variation in terminology; terminology standardised by expert committees would display the lowest degree of variation; terminology used in natural communication between specialists would have a medium degree of variation (our translation).

3. The study of the representation of variation and its causes

To study terminological variation, the pieces of information obtained from a complex approach to specialised discourse need to be taken into account. Delavigne (2017) grouped this information into four sections: (i) context (the most immediate linguistic context); (ii) conditions of production; (iii) conditions of circulation; and (iv) conditions of reception. However, it was Cabré (1998a) who went into the greatest detail about how these parameters affected and classified specialised discourse. Indeed, in 1998, she asserted that the dynamicity of specialised communication was what gave a sociofunctional value to specialised discourse. This discourse was defined by its conditions of production, starting with a certain conceptualisation, followed by the manifestation of the meaning of terminological units, and the conditions of reception, all of which determined how it was finally decoded.

Pecman (2014) also focused on discourse to understand the role of terminological variation in specialised communication. Basing her work on the Grammatical Metaphor Theory by Halliday⁴ (2004), she empirically showed that variation enabled specialists to make headway with the creation of new knowledge, through the reformulation of specialised concepts in discourse.

With the development of the concept of situated conceptualisation, Frame-Based Terminology (Faber et al. 2006; Faber, this volume) also provided a theoretical basis for the study of variation and for its textual analysis. In their most recent works,⁵ Tercedor (2013); León-Araúz (2017) and others have described a systematic methodology to record terminological variants based on cognitive and usage criteria, which allow these variants to be semantically and pragmatically characterised, and the cognitive processes of lexical variant comprehension and production to be studied. In their term base, denominative variants are accompanied by analyses that show the dimension of the concept highlighted in each denomination (see Table 1 for an illustrative example).

In the analysis of denominative variation, and with the aim of taking a closer look at its causes, there are other concepts that should also be taken into account such as whether different denominations are linked to one or various senders, i.e., a phenomenon of self-variation or of hetero-variation (Freixa 2006), since some causes are associated with either one or the other. Other issues that need to be taken

4. This observation is inspired by Halliday (2004), that is, a process of reformulating a concept through different grammatical categories in order to capture its essence and establish a new paradigm of knowledge. As such, grammatical metaphor would be the central feature of scientific discourse and the pathway that leads from experience to theory, and consequently to new knowledge (Pecman 2014, 2).

5. Produced within the framework of the VariMed project (<http://varimed.ugr.es>) Accessed 11 December 2021.

into account are whether variation by the same author occurs in one or several texts and whether variation (self-variation or hetero-variation) is intentional or spontaneous (Tercedor 2013).

Table 1. Dimensional variant types in the psychiatry corpus (León-Araúz 2017, 223)

Dimension	Term variant
+Discoverer	Korsakoff’s psychosis
+Symptom	burning-mouth syndrome
+Cause	alcohol-induced amnesic disorder
+Body_part	Broca’s aphasia
+Patient	boxer’s dementia
+Result	bedwetting
+Intensity	mild cognitive impairment
+Time	short-term insomnia
+Location	prison psychosis

It was from this perspective that Boschirolí (2013) analysed the variation in legal texts, namely, self-variation in *El Código Procesal Civil y Comercial de la Nación de la Argentina* [Code of Civil and Commercial Procedure of Argentina], in which *medidas cautelares* [precautionary measures] alternate with *providencias cautelares* [precautionary ordinances] and then with *medidas precautorias* [precautionary measures], where, in Spanish, *precautorias* [precautionary] is an alternative form of the adjective *cautelares* [precautionary].

Although, strictly speaking, these denominations correspond to different notions in the legal vocabulary, the author considers them to be denominative variants because of their contextual interchangeability. After analysing the same concept in other legal texts, she finds that denominations multiplied as a result of hetero-variation: *decisiones cautelares* [precautionary decisions], *cautelares* (where, in Spanish, the adjective *cautelares* [precautionary] is used as a plural noun) and, lastly, *cautela* [precaution]. The author concludes that whilst it was difficult to venture an explanation, it seemed to be a combination of stylistic constraints particular to the genre and individual stylistic concerns (Boschirolí 2013, 91).

From the perspective of Specialised Translation, Kerremans (2010) conducted a comparative study on terminological variation in different languages with the aim of enabling the future development of dictionaries reflecting intralinguistic and interlinguistic terminological variation, which would be of considerable help to specialised translation. Indeed, while terminographical products increasingly include denominative variation, what we see in the most representative terminological databases rarely goes beyond generic markers such as *see*, *see also*, *commonly known as*, *also known as*, *known as* and *also called* (Carrió and Candel 2012) or variant priority markers such as *preferred*, *admitted*, and *proposed*.

4. Classification of the causes of denominative variation

Freixa (2002) performed an exhaustive review of the literature that contained some reference to the causes of terminological variation. We found that there was a gap in this topic, as well as some confusion between the types and causes of variation. Based on these references and the research that we conducted in the field of Waste Management, we were able to offer a classification of the causes of denominative variation. This appeared in Freixa (2006) as shown in Table 2.

Table 2. Typology of causes of variation (Freixa 2006)

Type	Subtype
1. Preliminary causes	Linguistic redundancy Arbitrariness of the linguistic sign
2. Dialectal causes	Geographical variation Chronological variation Social variation
3. Functional causes	Adaptation to the level of language Adaptation to the level of specialization
4. Discursive causes	Avoiding repetition Linguistic economy Creativity, emphasis and expressiveness
5. Interlinguistic causes	Cohabitation of the 'local' term and the loanword Diversity of alternative proposals
6. Cognitive causes	Conceptual imprecision Ideological detachment Differences in conceptualization

In reality, there are only five blocks of causes because the first one refers to the linguistic factors favouring the existence of variation, and not strictly to the causes. Variation is a form of redundancy because languages are redundant, and the arbitrariness of the linguistic sign stems from the lack of a fixed link between denomination and referent, which allows a concept to be denominated in various ways. Thus, languages vary because variation is one of their intrinsic characteristics.⁶

Dialectal causes arise because of the different origins of the specialists that produce specialised texts. The difference in origin can be geographical (*diatopic variation*), temporal (*chronological* or *diachronic variation*) or social (*diastratic*

6. We agree with the nuance proposed by authors like Desmet, who use the term *variability* to distinguish it from *variation*: "Nous définissons la 'variabilité' comme la capacité de toute langue naturelle de produire de la variation lorsqu'elle s'actualise en discours." (Desmet 2006, 1) (We define 'variability' as the capacity that every natural language has to produce variation when it is updated in discourse) (our translation).

variation). The functional group of causes include variation produced by different specialised communication situations, and the discursive group includes variation due to stylistic and expressive factors. The sociolinguistic causes are associated with the contact between languages and, lastly, the cognitive causes, are linked to different conceptualisations as well as other possibilities.

Freixa (2013) argues that the causes are not mutually exclusive and that dialectal variation, for example, could reflect different conceptualisations. Even though more recent work has examined these causes in greater depth, there have not been any new classification proposals. In her work, Delavigne (2017, 34) states her interest in the following six types of variation (clearly referring to six different causes): diachronic variation, diatopic variation, diastratic variation, idiosyncratic variation, diaphasic or stylistic variation, and diamesic variation. Given that the first three types are dialectal and that the last two are functional, the novelty of the proposal lies in idiosyncratic variation, which is not strictly a cause, but rather a catch-all drawer for examples that seem to be governed by an individual or collective preference that cannot be attributed to another cause.

4.1 Discursive causes: Specialists and their texts

When specialists write a text, whether a highly specialised one for their peers or a mainstream one aimed at a non-specialised audience, they are bound by the rules governing the good drafting of any text. Thus, they must transfer specialised knowledge not only in an objective, precise, and systematic manner, but also in an expressive and rhetorical discourse. That is why they must, for example, avoid the constant repetition of those denominations referring to the most frequent concepts in their discourse. To do so, they must resort to resources such as variation.

On other occasions, specialists are confronted with the need to emphasise an aspect of the denominated concept, to shorten a long term or to show their creativity by finding other ways to denominate something. They must thus resolve such needs without creating a text that is either ambiguous or lacking in communicative effectiveness, owing to its expressive richness.

A good example is given in Fernández-Silva (2014), who analyses scientific texts in the field of Psychology by observing the denominative variation of certain concepts in different sections of the texts, according to the functions of each one: (i) title, abstract and keywords – identify and attract; (ii) introduction and conclusions – explain and generalise; (iii) methods and results – particularise. In this work, the author observed examples of denominative variation and identified their rhetorical function, which is understood to be any strategy that makes a text consistent and coherent. Thus, for example, after analysing the introduction (whose aim is to create a research space), she asserted the following:

En esta sección, la variación denominativa es funcional en el proceso de explicación del concepto, pues a través de distintas variantes se resaltan diferentes características del concepto y se ofrecen varios puntos de vista sobre su categorización (Fernández-Silva 2013). Como se muestra en los siguientes fragmentos de un estudio con adolescentes, el texto alterna la denominación *adolescencia* con otras variantes hiperonímicas que, al ubicar el concepto en clases conceptuales distintas, reflejan la diversidad de perspectivas desde las que puede ser abordado el concepto: desde la perspectiva del desarrollo biológico (*momento evolutivo*), como un periodo entre la infancia y la adultez (*etapa*), como un momento en la historia de vida (*época*) o como objeto de estudio de interés social (*tema de la adolescencia*).

(Fernández-Silva 2014, 55–56)⁷

As reflected in the example, the discursive need was resolved by the use of variants offering different types of information about the concept, a clear case of cognitive variation. This point of encounter between the rhetorical and cognitive functions had already been observed by Pecman (2014), where a concept was intentionally denominated in different ways to create an effect of newness and to emphasise the novelty of the scientific research presented in the text.

This does not mean that variation for discursive reasons correlates with variation for cognitive reasons. One of the most frequent mechanisms for avoiding denominative repetition is actually that of shortening, whether strictly contextual because of its anaphoric power, or lexical, when it can function autonomously without any conceptual ambiguity. Indeed, Lavagnino (2012) studied reductions in specialised texts and asserted that anaphoric variation had more of a contextual value since it strengthened the text's cohesion and consistency, whereas lexical reductions could give rise to true terminological variants, especially when core concepts of the text were concerned. The author illustrates her results with examples such as those in Table 3, which show the weight of each type of reduction.

7. In this section, denominative variation is functional in the process of explaining the concept because, by means of several variants, different characteristics of the concept are highlighted and various points of view with regard to its categorisation are offered (Fernández-Silva 2013). As shown in the following excerpts from a study relating to adolescents, the text alternates between the denomination *adolescencia* (adolescence) and other hypronymic variants that, by placing the concept into different conceptual classes, reflect the diversity of perspectives from which the concept can be approached: from the perspective of biological development (*momento evolutivo* [evolutionary moment]), as a period between childhood and adulthood (*etapa* [stage]), as a moment in the history of life (*época* [time]) or as a social-interest object of study (*tema de la adolescencia* [topic of adolescence]) (our translation).

Table 3. Polymorphy of complex terms in specialised texts (Lavagnino 2012, 50)

Full term	Variants
<i>mode de production biologique</i> (biological production method)	<i>mode</i> (16%) (method) <i>mode de production</i> (35%) [anaphoric reduction] (production method) <i>production biologique</i> (49%) [lexical reduction] (biological production)

4.2 Dialectal causes: Diatopic, diastratic and diachronic variation

This section addresses the dialectal causes of denominative variation, which result from the fact that people belong to a group, which can be geographical, chronological or social. It reviews the work done in recent years and the extent to which they provide new information regarding the three types of dialectal cause: diatopic, diastratic and diachronic variation.

4.2.1 *Diatopic variation*

The variation produced when different geographical zones where the same language is spoken opt for different denominative solutions is called *geographical variation*. However, this type of variation can also be called *geolectal variation*, *diatopic variation* or *topolectal variation*, thus producing a good example of ‘metavariation’.

It is the most obvious and widely studied type of variation, to the extent that Wüster acknowledged it in his General Theory of Terminology, as do his followers. However, what they condemned was the risk of ambiguity that it, like any other type of variation, might entail. It is indeed an undeniable phenomenon, although it does not affect all languages or knowledge areas in the same way. For example, diatopic differences are more pronounced in spoken languages in different countries, especially if they are farther away from one another (either physically or culturally), and the knowledge areas most affected by variation are those that are closer to the culture and activities of everyday life.

For example, Bomrad (2012) describes in great depth the terminological variation of grapevines and wine by comparing the Spanish of the Iberian Peninsula to that of Argentina. She provides numerous examples like the following:⁸ *cosecha tardía* / *vendimia tardía* [late harvest, in both instances], *desborre estático* / *desfangado estático* [static racking, in both instances], *espaldero* / *espaldera* [espalier, in both instances], *parral* / *emparrado* [vine training, in both instances]. Bestué (2016)

8. The first denomination refers to the Spanish of Argentina and the second to that of Spain.

analyses the intralinguistic variation produced by geographical causes in an area highly related to society as a whole, namely, its legal system. In this regard, she gives the following example:

El “derecho de sociedades”, “derecho de corporaciones” o “derecho corporativo”, es aquella disciplina que estudia las formas sociales, tanto civiles como mercantiles. Estas tres denominaciones responden fundamentalmente a la variación terminológica derivada de su adscripción a un determinado ordenamiento jurídico.

(Bestué 2016, 20)⁹

Bernal (2017) describes an interesting example of geographical terminological variation that has occurred in recent years because of the increasing numbers of people using bicycles as a means of transport, which has led to the proliferation of cycle lanes on the public highways of many towns and cities. Spanish-speaking countries have resolved this denominative need by using different terms: in Spain, a cycle lane is called a *carril bici*; in Argentina it is a *bicisenda*; in Mexico, they have opted for *ciclopista*; in Peru, for *ciclovía*; and *ciclorruta* is the term shared by Chile, Colombia and other countries.

4.2.2 *Diastratic variation*

Affiliation to a social group, i.e., social, sociolectal or diastratic variation, is a cause of variation that has been studied from the perspective of Sociolinguistics for decades. In the case of specialised discourse, it is affiliation to a certain socioprofessional group rather than to a social group that can generate denominative variation in the various fields of knowledge. This aspect of variation was highlighted and studied by various authors who are advocates of *Socioterminology* (Gaudin 2003), which confers an important role on the observation of the conditions of specialised knowledge production, on the social practices where such production took place, and the conditions of the utterance in general.

Quirion (2013) analyses the surveys of terminology implantation in Quebec 15 years after the adoption of the *Charte de la langue française* (Charter of the French Language) in 1979, and he identifies the decisive factors when it came to using one of three terminological options: the English term, the popular term, or the French term. The results show that the cause that clearly stood out to those conducting the survey concerned people's membership in a social or socioprofessional group (Quirion 2013, 49). His analysis further specifies that the level occupied within a

9. ‘Derecho de sociedades’ (company law), ‘derecho de corporaciones’ (corporation law) or ‘derecho corporativo’ (corporate law) is the discipline that studies corporate forms, both civil and commercial. These three denominations basically respond to the terminological variation derived from their affiliation to a given legal system (our translation).

business hierarchy was a determining factor for the denominative choice, and that the desire to join or remain in a group especially manifested itself in spoken use.

Picton and Dury (2017) also studied diastratic variation in specialised texts, specifically in the areas of nuclear medicine and university teaching. Their analysis compared different professional groups within the same sector and significant differences were observed. For example, they found denominative differences between nuclear medicine doctors and technologists because the approach of nuclear medicine doctors was more theoretical and the approach of technologists was more applied. This was also observed in the university teaching sector, where, for example, *dossier d'enseignement* [teaching dossier] alternated with *portfolio d'enseignement* [teaching portfolio] in administrative texts, whereas the second denomination was the only one used in academic articles. The authors further note that variation between a full term and its apocope was one of the most common diastratic variations in the field of Nuclear Medicine: *cardio/cardiologue* [cardio/cardiologist], *tomo /tomographie* [tomo scan/tomography scan], *perf / perfusions* [IV / intravenous therapies], *vent /ventilation* [vent /ventilation], and *acqui / acquisition* [acqui / acquisition], among many others.

The examples given so far (excluding the apocopes in the preceding paragraph) show that as previously observed, the causes of variation are not mutually exclusive. The denominations used by different professional groups may be the result of different conceptualisations, a phenomenon that will be analysed in the discussion of cognitive causes. Similarly, geographical variation may originate from different approaches to a concept and, ultimately, different conceptualisations. Furthermore, the discursive variations that we saw previously could be analysed from a cognitive viewpoint.

4.2.3 *Diachronic variation*

New approaches to Terminology incorporate the study of diachrony into research lines in order to examine its contribution to our knowledge of specialised discourse and to the behaviour of terms. Work carried out from a diachronic perspective incorporated quantitative analyses, includes Nazar (2011) and Picton (2009). Nazar (2011) studied the evolution of Terminology in the field of Linguistics in recent years by analysing frequencies of use, and the Picton (2009) proposes a tool for analysing the evolution of knowledge in specialised corpus in the field of Aerospace.

One of the most productive aspects of the diachronic study of terminology is the denominative variation brought about by the evolution of specialised knowledge. This implies the coexistence of an established term alongside a new term to denominate a concept. This occurs either because the concept has changed in some way or because the established denomination is no longer deemed valid by

the specialists. However, denominative variation can also arise from the birth of new concepts. This phenomenon had already been observed by Guilbert (1973), who asserts that vocabularies require a period of development (he called them *vocabulaires en gestation* [vocabularies in gestation]) characterised by a low level of terminological fixation and the emergence of *variation transitoire*, or transient variation, which is resolved over time through conceptual stabilisation and the consequent denominative stabilisation.

Dury has analysed diachronic variation in various specialised fields (textile industry, medicine, and ecology, among others). Dury and Lervad (2008) include a good example of the transient diachronic variation mentioned in Guilbert (1973). When the authors traced the concept of PETROLEUM in texts from different time periods, they found concurrent terms with poorly defined semantic contours because they referred to little known and inadequately described concepts that were likewise poorly defined:

[...] les synonymes décrivent en effet tantôt l'origine du produit, autrement dit où on le trouve dans la nature (*rock oil*, *mineral pitch*, *earth oil*), ce qu'il sent (*empyreumatic oil*), quel aspect il a (*liquid bitumen*, *dark pitch*, *tar*, *carbon oil*) et aussi dans quel pays ou quelle région on le trouve. [...] Tous ces termes constituent ce que l'on a appelé la *variation synonymique de circonstance* autour de *petroleum*, et n'appartiennent en fait, pour la plupart, que de façon provisoire au lexique : en effet, la totalité des métonymes toponymiques ont aujourd'hui disparu de la terminologie du pétrole, ainsi que la plus grande partie des termes décrivant l'origine minéralogique du produit. Peu de termes sont restés, comme *naphtha*, *asphalt*, *bitumen*, *maltha*, mais tous ont vu leur sens se spécialiser pour continuer d'exister. Ainsi le terme *naphtha* désigne t'il aujourd'hui un produit utilisé comme combustible, et obtenu par distillation du pétrole ou d'autres corps riches en matières organiques. (Dury and Lervad 2008, 72–73)¹⁰

10. [...] the synonyms actually describe the origin of the product, i.e., where it is found in nature (*rock oil*, *mineral pitch*, *earth oil*), as well as its smell (*empyreumatic oil*), its appearance (*liquid bitumen*, *dark pitch*, *tar*, *carbon oil*) and the country or region where it can be found. [...] All of these terms constitute *circumstantial synonymic variation* – as it has been called – around *petroleum* and, for the most part, they only belong to the lexicon in a temporary way: in fact, all of the toponymical metonyms have now disappeared, as have most of the terms describing the mineralogical origin of the product. There are very few remaining terms, such as *naphtha*, *asphalt*, *bitumen* and *maltha*, though the meaning of each one has become more specialised in order to continue to exist. Thus, the term *naphtha* now designates a product used as a fuel, which is obtained by distilling petroleum or other bodies rich in organic materials (our translation).

Hence, this type of circumstantial synonymic variation does not remain fixed in the terminology of a field and disappears in subsequent stages or becomes semantically specialised (i.e., it goes on to cover a more specific specialised concept). Moreover, the different denominations in existence at a given moment in time show the different conceptualisations of the same concept, as mentioned in Freixa (2013, 43):¹¹

Las causas previas y las causas cognitivas de hecho son el origen de la variabilidad y explican las alternancias denominativas provocadas por las causas dialectales, funcionales, estilísticas y las interlingüísticas (cuando hay alternancia en las denominaciones para desplazar un préstamo).¹²

Dury has reflected extensively on the denominative need that specialists feel they have, especially in the health area, where connotation is decisive. This need leads them to create new denominations when, with the passing of time, the existing ones acquire undesired connotations or lose desired ones. That makes it a source of denominative variation, which is usually resolved by the disuse of the undesired denomination (i.e., *terminological obsolescence*) and the implantation of a neological term. Thus, for example, *health anxiety* and *severe health anxiety* are the preferred neological forms for *hypochondriasis*; *cardiovascular malformation* and *congenital heart anomaly* are preferred to *congenital heart defect*; and *minimally conscious state* is preferred to *vegetative state* (Dury 2012, 87).

Parizot (2014) also took a socioterminological approach and conducted an ethnographic study in the field of Business, where she observed that numerous denominations had changed diachronically for reasons of connotation. She gives a very telling example of the evolution from *voyageur* [travelling salesperson] to today's *responsable de comptes secteur* [sector account manager], with *responsable technico-commercial* [technical sales manager] and other denominations like *représentant* [representative] and *commercial terrain* [field salesperson] in-between:

11. And Dury (2013, 3) likewise states: “Bien que cette classification intègre la variation temporelle dans la variation dialectale, nous avons toujours favorisé l'idée (...), qu'elle est aussi fortement dépendante des variations qui peuvent se produire, au fil du temps, au niveau conceptuel et qu'elle fait donc également partie intégrante de la variation cognitive. Nous reviendrons sur ce point plus tard au sujet de la connotation et de l'obsolescence des termes.” (While this classification integrates temporary variation into dialectal variation, we have always favoured the idea [...] that it is also highly dependent on variations that can occur over time at the conceptual level, and that it is therefore an integral part of cognitive variation too. We will return to this point later, when discussing connotation and the obsolescence of terms (our translation)).

12. In fact, prior causes and cognitive causes are the origins of variability, and they explain the denominative alternants brought about by dialectal, functional, stylistic and interlinguistic causes (when there is alternation in the denominations to displace a loanword) (our translation).

Les commerciaux sur la route se sont appelés jusque dans les années 1980 *voyageurs*. Ils étaient les représentants de l'entreprise auprès de la clientèle. Puis l'entreprise a utilisé *responsable technico-commercial* (RTC). Cependant pour des raisons de connotation négative liée au déterminant *technico-commercial*, celui-ci a été abandonné. Il s'est transformé pour devenir *responsable de comptes secteur* (RCS). Dans les parlers de l'entreprise, le terme *représentant* est encore employé comme synonyme, car il correspond au concept valorisant à ce titre de représentation de l'entreprise. Il est cependant moins courant que *commercial terrain* qui lui correspond. (Parizot 2014, 53)¹³

The variation between these denominations once again highlights conceptual plasticity and shows how changes in circumstances, interests, and needs over time are resolved with new denominations that place greater emphasis on one dimension of the concept than another.

4.3 Functional causes

Within the framework of the CTT, Cabré (1998a) contends that functional variation parameters (channel, field, tenor, and tone) are also applicable to the study of variation in specialised communication. Regarding tenor and tone, Cabré asserted that a communication's degree of specialisation determines not only the terminological density of a text, but also the quantity of expressive variation for referring to the same concept. Thus, a highly specialised text is normally more precise, concise and systematic than a less specialised one. This hypothesis was the one that we explored in Freixa (2002). In that work, we compared denominative variation in waste recycling texts with varying degrees of specialisation, and found that the amount of variation was indeed greater in less specialised texts. In fact, we identified three major differences in variation between texts by their respective degrees of specialisation: (i) the lower a text's degree of specialisation, the higher the percentage of concepts represented by more than one denomination; (ii) the lower a text's degree of specialisation, the higher the quantity of different denominations for the same concept; (iii) the lower a text's degree of specialisation, the lower the semantic equivalence between denominations.

13. Up to the 1980s, salespeople on the road were called *voyageurs*. They were the client-facing company representatives. Then the company started using *responsable technico-commercial* (RTC). But, for reasons of negative connotation linked to the determiner *technico-commercial*, this title was dropped. It became *responsable de comptes secteur* (RCS). In the company's business-speak, the term *représentant* is still used as a synonym because it corresponds to the concept of representing the company. However, it is less common than *commercial terrain* (our translation).

The example in Table 4 shows how the same concept is denominated by two terms in texts with a higher degree of specialisation (*biodegradación* [biodegradation] and *descomposición* [decomposition]), whereas it is given six different denominations in texts with a lower degree of specialisation. Regarding the latter, some are more paraphrastic than others, the purpose of which is to ensure that the concept can be understood.

Table 4. Comparison of variation for *descomposición* (Freixa 2002)

Corpus < specialised	Corpus > specialised
<i>descomposición</i> (<i>decomposition</i>)	<i>biodegradación</i> (<i>biodegradation</i>)
<i>proceso de descomposición</i> (<i>decomposition process</i>)	<i>descomposición</i> (<i>decomposition</i>)
<i>descomposición biológica</i> (<i>biological decomposition</i>)	
<i>degradación</i> (<i>degradation</i>)	
<i>proceso de degradación</i> (<i>degradation process</i>)	
<i>proceso natural de degradación</i> (<i>natural degradation process</i>)	

Based on corpus analysis, Carrió and Candel (2012) study terminological variation from a pragmatic perspective. Their results confirm the important role played by adaptation to the audience in specialised communications:

We observed in this analysis that terms are not used in a random way by technical writers. The use of the terms depends on the pragmatic perception of the writer, who adapts a text for a particular target audience. [...] In this paper, we used a corpus-based study to show that term variation occurs depending on the pragmatic implications of the text, as Bowker and Hawkins (2006) explained; but also, the intention of the writer and the audience are important in order to decide whether a standard term or a variant is more appropriate. [...] In this study, we consider that social and conceptual motivations are the most frequent when choosing variants. In this sense, the pragmatic implications derived from our analysis are evident. The sender and receiver of the message play a vital role in term variation.

(Carrió and Candel 2012, 48)

In reality, adaptation to the receiver’s level of specialisation is the most widely studied functional cause of denominative variation, whereas the channel is the least analysed one. Bowker and Hawkins (2006, 98) mention a clear terminological difference between spoken and written medical texts, though they did not provide any empirical data. Seghezzi (2013) provides such data and compares denominative

variation in spoken and written texts in the field of Linguistics (lectures and presentations, and published texts of spoken texts, respectively). Even though she did not find any significant differences between spoken and written texts at a quantitative level, she did find them at a formal level because the written denominations were more lexicalised, elaborate, and complex than oral denominations, which were more analytical, transparent, and audience-friendly. Furthermore, she observed a discursive-textual contrast because the function of terminological variation in the written texts was, to a large extent, stylistically motivated in order to achieve a more elaborate discourse, whereas a cognitive function of terminological variation was evident in oral texts to facilitate listeners understanding of the concepts (Seghezzi 2013, 74).

4.4 Sociolinguistic causes: Contact between languages

Freixa (2006) asserts that the contact between one language and others is one of the causes of denominative variation that had received most attention, and research interest in this cause has continued to the present. Contact between languages is not only understood as contact due to geographical contiguity, but also to cultural contiguity. Studies almost always refer to contact between English – as the language for exchanging specialised knowledge within the international community – and other languages. The use of English terms in other languages, which we call loanwords, is not only a matter of prestige, but also evidence of the fact that the English term is communicatively more efficient, since it is generally the most widely known in different languages. However, it also demonstrates how difficult it is to change the denomination when the speakers have already incorporated the loanword into their vocabularies.

In Spain's newspapers, for example, the use of the loanwords, *outsourcing*, *router*, *wearable*, *share*, *subprime* and *gamer* is much more frequent than their corresponding Spanish denominations *tercerización*, *direccionador*, *dispositivo vestible*, *cuota de pantalla*, *hipoteca de alto riesgo* and *videojugador*. In French, Dury and Lervad (2008) studied examples such as *lissage* (smoothing) and *remodelation* (remodelling), which had not managed to become successfully implanted because the specialists considered *lifting* to be a more technical term.

Contact between languages is even more significant in some instances when, besides coming into contact with English, one language is pressured by another language. This is the case, for example, of Catalan, a medium-sized language that is trying to develop itself terminologically in the shadow of not only two strong languages such as Spanish and English (Freixa 2019), but also of many other languages including Arabic. Indeed, in North Africa, the conflict of prestige between

Arabic and the two colonial languages (French and English) also gives rise to a large quantity of denominative variation (Hamed 2016).

Some studies have analysed another source of variation associated with contact between languages, namely the emergence of different solutions to try and avoid the use of loanwords. Casals (2006) illustrates this situation with the translation into Spanish of the English term *delusion* in the field of Psychiatry, where the Spanish terms, *delirio* [delirium], *idea delirante* [delirious idea] and the adapted loanword *delusión* [delusion] all coexist. The author emphasises the ‘emboldened’ translator role of specialists:

La variación terminológica se incrementa debido a la facilidad y rapidez de comunicación con la que contamos. Un número cada vez más alto de científicos lee en inglés y, a la hora de expresarse en español, « arriesgan » una traducción del término mediante una simple fonologización del término en lengua original (normalmente inglés) o una creación a menudo afectada por el posible desconocimiento de la lengua materna de los científicos. De esta manera, los consensos a los que se llegan los elaboran los miembros de comunidades muy reducidas, en lugar de acordarlos entre todos los expertos del mundo hispanohablante que puedan usar dicho término. Y este consenso general sobre la correspondencia entre término y concepto hace vulnerable, o inexistente desde un principio, la precisión terminológica, ya que el significado del término varía desde un principio según el contexto. (Casals 2006, 223)¹⁴

4.5 Cognitive causes

The various terminological approaches that currently exist acknowledge the dynamicity of knowledge, the flexibility of conceptual structures and the diversity in object conceptualisation. Denominative variation enables us to observe the different dimensions of a concept or the different viewpoints from which it can be approached.

14. Terminological variation is increasing due to the ease and speed of communication we enjoy today. More and more scientists read in English and, when it comes to expressing themselves in Spanish, they make ‘bold’ attempts to translate a term by means of a simple phonologisation of the term in the original language (usually English) or a creation that is often affected by a potential lack of knowledge of the scientists’ mother tongue. Thus, consensus is developed and reached by the members of very small communities rather than the Spanish-speaking world’s experts as a whole, who are the ones likely to use the term. And such consensus on the correspondence between a term and a concept means that terminological precision is either vulnerable or non-existent from the outset, since the meaning of a term varies from the very beginning according to the context (our translation).

The cognitive causes of variation have undoubtedly been the most widely studied in the past decade, and various authors highlight the different motivations, visions, intentions and connotations of the forms denominating the same concept. And the fact is that different interpretations of reality may either be spontaneous or correspond to an intention or interest. Different professional groups or different schools of thought, for example, may naturally approach a specialised concept in different ways, but such diversity may also be due to deliberate distancing decisions.

Whatever the case may be, Bowker (1998, 488–489) explains that variation between specialists is not arbitrary, and she defends the weight of multidimensionality in this phenomenon:

We use the term *multidimensionality* to describe the phenomenon of classification that occurs when *more than one* characteristic can be used to distinguish between things, and hence those things can be classified in more than one way. A *dimension* represents one particular way of classifying a group of things; a classification with more than one dimension is said to be *multidimensional*.

Shortly afterward, Temmerman (2000) relates variation to the possibility of showing different perspectives or viewpoints by using different denominations. This is illustrated with the now classic example of a genetic engineering technique developed by E. M. Southern for which there are several denominations (*Southern blotting*, *Southern transfer* and *Southern hybridisation*). These variants reflected the aspect highlighted (the result, the process or the principle). Temmerman (2000, 151) asserts that such variation contributes to better communication:

The naming of Southern's technique is motivated by (at least) three aspects of the (encyclopaedic) information concerning the technique [...]. Depending on which of the three elements a language user wants to stress in a particular context the use of one or another of the synonyms can be appropriate for more nuanced communication.

Fernández-Silva (2011, 2013, this volume) has studied denominative variation most extensively from a cognitive perspective within the framework of the CTT. In her 2011 work, she addresses the existence of two types of denominative variation: one that involves cognitive variation and another that does not. This distinction has been fundamental to subsequent studies, and her research not only studies the viewpoint, which she understood as the position from which a certain author approaches a certain concept, which has an influence on the perception of what the concept's distinctive characteristics are (Fernández-Silva 2013, 12), but also to investigating other aspects that have enabled a better understanding of the cognitive role in terminological variation.

Frame-Based Terminology (Faber et al. 2006, this volume) approaches terminology from a cognitive perspective, and interesting works on terminological variation have been produced. In one way or another, such works have led to the analysis of cognitive causes:

Desde el punto de vista pragmático y cognitivo, las formas de conceptualizar una misma realidad que se recogen bajo el paraguas de la variación denominativa incluyen unidades que lejos de poder entenderse como sinónimos totales o parciales, revelan motivaciones situadas derivadas del contexto, entendido éste en sentido amplio y dinámico. (Tercedor 2013, 84)¹⁵

Within cognitive causes, some authors have highlighted the fact that denominative variation also arises from a desire for ideological distancing from the conceptualisation embedded in a certain denomination. We saw this in the analysis of diachronic variation, where specialists preferred to reject a denomination because it was perceived as having connotations, or simply because it reflected a superseded stage of knowledge:

Ce besoin de distanciation idéologique peut aussi se produire lorsque le terme correspond à des cadres théoriques aujourd'hui considérés comme désuets, appartenant à des générations antérieures. Dans ce cas, le caractère péjoratif du terme, et son rejet par la communauté scientifique, peut venir de ce qu'il exprime des paradigmes scientifiques désormais abandonnés. (Dury and Lervad 2008, 7)¹⁶

Palomero (2020) conducted a study on the denominative variation of the concept of *violencia de género* [gender-based violence], in which, besides controlling for the time variable, he examined how the sender's ideology and attitude were reflected in the choice of one denominative variant or another. He found that *malos tratos* [abuse], a denomination that hides the violence and the victim, had gone from being the most frequently used one in the 1990s to one that is only occasionally used today. In contrast, the opposite had occurred with *violencia de género*, the more legal-sounding and more neutral denomination, which is now the most frequently used. Other denominations, such as *violencia machista* [male violence] and *violencia contra las mujeres* [violence against women], which are clearly not

15. From a pragmatic and cognitive viewpoint, the forms of conceptualising the same reality, which are grouped under the umbrella of denominative variation, include units that, far from being able to be understood as full or partial synonyms, reveal situated motivations derived from the context, the latter being understood in a broad, dynamic sense (our translation).

16. This need for ideological distancing may also arise when the term corresponds to theoretical frameworks that are now deemed outdated, belonging to previous generations. In this case, the pejorative character of a term, and its rejection by the scientific community, may come from what it expresses about scientific paradigms that have now been abandoned (our translation).

neutral because they denounce the aggressor and make the victim explicit, are closer to a progressive ideology with a certain feminist vindication, whereas *violencia doméstica* [domestic violence] and *violencia intrafamiliar* [intrafamily violence], which put denominative emphasis on where the violence takes place, are generally used by groups that do not even accept the existence of male violence against women.

To end the review of the cognitive causes of denominative variation, there are authors who, in the culturally different views of the world, have observed yet another example of cognitive variation. Faber and Medina-Rull (2017) explore the cultural dimension in the categorisation of WIND and observe that its classification and denomination are derived from the cultural perception of the phenomenon.

Worthy of special mention is Diki-Kidiri (1999, this volume) with his cultural approach to Terminology. Based on his study of oral-tradition African languages, the cultural component in the perception, categorisation, and denomination of reality is clear. He provides different examples of variation and asserts that the most suitable and most widely accepted denomination is very often the one that fits best into the community of speakers' language and culture (Diki-Kidiri 1999, 580).

5. Conclusions

While studies on the causes of denominative variation conducted in the past decade were the main focus of this non-exhaustive literature review, a small number of earlier ones have been examined in this chapter. The review demonstrates that the study of variation is at the heart of recent lines of terminological research because all new proposals take an interest in it. The study of the causes of denominative variation also figures prominently within that research because it is crucial to provide a more in-depth explanation of how and why speakers resort to variation in different specialised communication situations.

There has been some debate on the status of the causes. Having overcome the confusion between causes and types, we observed that it had gone on to affect the functions. On the one hand, the types of variation generally refer to groups by form (lexical variation, morphosyntactic, orthographical, etc.) and the causes refer to the reasons why such variations exist, and to their origins. On the other hand, the function of a particular type of variation pertains to the role it plays and the purpose it serves. This distinction is much clearer on some occasions than it is on others. Within the discursive causes, for example, we could say that true functions include avoiding repetition and achieving a degree of linguistic economy as well as a greater emphasis or creativity. The cause is the accomplishment of the function.

The literature review in this chapter broadly corroborates the relevance of our 2006 proposal of the causes of variation, though some work does incorporate new proposals. Even though they have yet to be fully studied, once they have been examined in greater depth, they may enhance and complete the initial proposal. We are referring to research such as Condamines (2017), who explores the emotional dimension of the variation of terms. Emotions play an important role in Cognitive Linguistics because they intervene in our perception and, therefore, in our knowledge. Her examples show that emotional perceptions shared by a speech community should be taken into account. In fact, Dury and Lervad (2008) also refer to the sender's commitment to a variant and the receiver's emotional response.

Finally, the examples given by different authors confirm that the causes of denominative variation are not all at the same level. As proposed in Freixa (2013), it is necessary to break with the idea of linearity and see groups of causes at different levels. We suggested that prior causes should be a separate set because they enable the existence of variation. Similarly, cognitive causes need to be treated differently because they explain the denominative alternants arising from dialectal, functional, stylistic and, occasionally, interlinguistic causes. Thus, when confronted with any case of denominative variation, the analysis can be performed at three different levels and from three different viewpoints: first, by the types of formal change between the denominations; second, by the conceptual differences between the denominations; and, lastly, by the cause of the alternant.

Diachronic variation

Pascaline Dury

Université Lumière Lyon 2

The use of the diachronic perspective to study specialised terms and concepts is a research area which has long been under-exploited in Terminology. Until very recently, studies on specialised terminologies were carried out only from a synchronic perspective, which is still predominant, although the use of the diachronic perspective is now less marginal. Specialised diachrony, i.e., the many changes occurring in terms and concepts over time should be considered as an integral part of term variation. Moreover, it gives access to terminological knowledge which is not often described when studied in synchrony, and it also provides knowledge about the conceptual structuring of specialised fields.

Keywords: specialised diachrony, diachronic variation, diachronic corpus, synchrony, neology, term obsolescence, term connotation

1. The slow emergence of an interest in diachronic variation

The inclusion of an entire chapter on diachronic variation in a volume on theoretical approaches to Terminology would probably have been considered completely irrelevant only a few decades ago, as the diachronic perspective was then of only marginal interest in the field. This shows how much the theoretical principles of Terminology, as well as research practices, have evolved since then. This chapter provides an overview of diachronic variation in Terminology, how it has slowly emerged as a research topic and how it has found its place in the recent work of terminologists. It will also show that diachronic variation has been an integral part of the general evolution of Terminology, which is described by Faber (2009, 17) as the shift “from prescriptive to descriptive with growing focus on the study of specialized language units from a social, linguistic and cognitive perspective”.

1.1 The predominance of the synchronic perspective in early research work

Even if the study of diachronic variation seems less marginal today in the research articles published by terminologists – both from a theoretical and methodological point of view –, it was at first largely ignored by them. As Picton (2009, 1) observes, Terminology was from the outset locked in a synchronic perspective that excluded any observation of the evolution of specialised terms and concepts over time.

D'un point de vue théorique, la terminologie a en effet longtemps été dominée par la perspective classique wüsterienne, dont l'objectif premier de normalisation exclut la variation temporelle, autrement dit la dimension diachronique ; cela explique que, par conséquent, d'un point de vue technique, très peu de ressources et d'outils existent aujourd'hui pour aborder cette perspective.¹

The few studies on Diachronic Terminology published in the 1980s and even 1990s by researchers such as De Schaetzen (1989), Møller (1994, 1998), and Van Campenhout (1998), were not part of a more general movement of interest in the diachronic perspective. For a long time, the terminological research studies carried out on the diachronic dimension were considered to be on the fringes of the field. The first and perhaps the most important reason for this is related to Wüster (1976, 51), who, from the outset, established the synchronic approach as one of the three main principles of Terminology:

Première particularité : le domaine des notions est considéré dans la terminologie, indépendamment du domaine des termes. [...]

Deuxième particularité : pour les terminologues, seul le vocabulaire est important. La morphologie et la syntaxe peuvent être empruntées à la langue courante.

Troisième particularité : la considération terminologique d'une langue est fortement synchronique ; pour la terminologie, le plus important dans une langue, c'est le système de notions qui se trouve à la base.²

1. From a theoretical point of view, Terminology has indeed long been dominated by the classical Wüsterian perspective, whose primary objective of standardisation excludes temporal variation, in other words the diachronic dimension; this explains why, from a technical point of view, very few resources and tools exist today to address this perspective (our translation).

2. First feature: The domain of notions is considered in Terminology independently of the domain of terms [...]. Second feature: For terminologists, only the vocabulary is important. Morphology and syntax can be borrowed from everyday language. Third feature: The terminological consideration of a language is strongly synchronized; for Terminology, the most important thing in a language is the system of notions that lies at the base (our translation).

Even though, as observed by Candel (2004),³ Wüster's writings show that he acknowledged diachronic variation, he did not consider it to be part of Terminology, which was therefore defined as a purely synchronic object of study.

The diachronic perspective was also ignored in the first books and articles on Terminology, whether they dealt with theoretical or methodological issues. Moreover, when authors used a diachronic perspective in their works, most of them did so only in the context of the study of new terms. The study of terminological evolution was thus strictly limited to the study of neology. Gilbert (1977, 65) writes for instance that:

Un examen du comportement des néologismes scientifiques et techniques dans le temps suppose que l'on admet au moins les trois préalables suivants [...] :

- a. On ne met pas en cause le caractère opératoire du concept de « néologisme » [...].
- b. On reconnaît la « spécificité du terme scientifique et technique » analysé par Louis Guilbert (1973) [...].
- c. On se place dans une perspective diachronique [...], ce qui va nous amener à envisager, par un découpage arbitraire, plusieurs phases dans la « vie » des néologismes en cause.⁴

Similarly, Dubuc (1992, 117) mentions diachronic changes in specialised lexicons only in relation to what he calls “technological neology”, a term he uses to designate any new reality, such as a new machine that is put on the market, or a new disease that has just been identified.

As observed by Dury and Picton (2009, 32), of the founding principles of Wüster that were later challenged by linguists, it seems that the synchronic study of the language was the least controversial and was therefore the last one to be questioned:

3. According to Candel (2004, 24), the theoretical and methodological predominance of synchrony in Wüster's work must be placed in the context of the linguistic research of the time, which favoured the study of language in an essentially historical perspective. The choice of synchrony as one of the founding principles of terminology was therefore, paradoxically, quite innovative at the time.

4. An examination of the behavior of scientific and technical neologisms over time is based on at least the following three prerequisites [...]:

- a. The operative character of the concept of “neologism” is not called into question [...].
- b. The “specificity of the scientific and technical term” analysed by Guilbert (1973) is accepted.
- c. We place ourselves in a diachronic perspective [...], which will lead us to consider, through an arbitrary division, several stages in the “life” of the neologisms in question (our translation).

Une partie des principes de base wüsteriens ont été remis en cause (la monoréférentialité, l'univocité, l'absence de connotation, la fixité des domaines par exemple), mais le principe d'un traitement synchronique strict est peut-être de tous les postulats de départ celui qui a été le moins remis en question.⁵

The diachronic dimension has consequently been the victim of a double neglect in Terminology, because of its absence in the first classical works of the field as well as in the work of terminologists who subsequently questioned the founding premises of the discipline.

Another, more recent obstacle to the development of the diachronic perspective is undoubtedly linked to the fact that very few computer tools specifically designed for the exploration of diachronic corpora have been developed in Terminology or have been made widely available to the research community. In most cases, terminologists working in diachrony and who are not fortunate enough to have dedicated, in-house tools at their disposal, must explore diachronic data using programs originally designed for synchronic analysis.

Another reason is probably related to the term *diachrony* itself, which is often misinterpreted, even by terminologists, some of whom regard *diachrony* and *history* as synonymous, and who consider that the study of terms and concepts from a diachronic perspective inevitably involves exploring the history of the language, over long or very long periods of time in order to describe structural changes affecting the language as a whole. Strangely enough, the study of changes and micro-changes that occur at the lexical level over a period of say, 5 to 10 years, is usually not considered to be part of the diachronic perspective.

1.2 The late legitimization of the diachronic perspective

Despite this general lack of attention some pioneering work on Diachronic Terminology was carried out in the late 1980s and 1990s, such as the proceedings of the first conference on the diachronic perspective in Terminology, held in Brussels in 1988 and published by De Schaetzen (1989). The papers presented describe case studies based on the diachronic perspective, and although none of them offer theoretical reflections on the role and place of the diachronic perspective in specialised languages, most of them deal with research issues that are still relevant today. This is the case with the relationship between neology and diachrony, for example, or

5. [...] Some of the basic Wüsterian principles have been called into question (e.g., mono-referentiality, univocity, absence of connotation, fixedness of domains), but the principle of strict synchronicity is perhaps the least discussed of all the founding postulates [...] (our translation).

the way in which conceptual organisation and reorganisations in a new specialised field are reflected in the language used by experts (see Goffin 1989).

Two papers by Møller (1994, 1998) are also frequently cited in research on Diachronic Terminology, as both combine theoretical reflection and descriptive analysis, which was quite new at the time. Furthermore, both of these papers were based on corpus analysis, which was also very innovative. They described research issues which were then rarely mentioned in Terminology, such as the obsolescence of terms or the different structures or levels of language at which diachrony can be explored (i.e., the micro-level at the level of terms and the macro-level at the level of terminology).

Much later, Candel and Gaudin (2006) published a collection of research works focusing on the diachronic aspects of specialised vocabulary. This publication marks a turning point in the field of Diachronic Terminology since it is one of the first one to present the diachronic perspective as being an integral part of terminological variation. It is also one of the first to highlight the need to take context into account in diachronic studies, something that has since become one of the starting points when analysing diachronic corpora. In the preface of the volume, the editors regret the lack of interest in Diachronic Terminology and hope that “les contributions réunies ici permettront d’attirer l’attention sur un champ d’études situé à l’écart des grandes voies fréquentées par les chercheurs en sciences du langage.”⁶ (Candel and Gaudin 2006, 13)

The new theoretical currents that have emerged in Terminology since the 1990s have also greatly contributed to the integration of the diachronic perspective into the main lines of terminological research and have played a major role in bringing it out of the marginal position in which it had hitherto been confined. Socioterminologists, particularly, have been among the first ones to call for a study of terms in their social dimension and for taking into consideration their evolution over time. According to Gaudin (2003, 38):

Il faut se déprendre d’une vision qui assimile termes et stabilité. [...] ce qui fonde la spécificité des termes, leur caractérisation par des valeurs exactes, oblige à considérer leur historicité car ces valeurs exactes attachées aux termes sont provisoires et révisables. Il faut se déprendre d’une vision qui assimile termes et stabilité.⁷

6. [...] the articles collected [...] will help to draw attention to a field of study that is far removed from the main paths followed by [...] researchers (our translation).

7. We must move away from a vision that equates terms and stability. [...] The specificity of the terms, their characterisation by exact values, makes it necessary to consider their historicity because the exact values attached to the terms are provisional and revisable. We must depart from a vision that assimilates terms and stability (our translation).

Within the framework of Sociocognitive Terminology, Temmerman (2000, see also chapter in this volume), drawing on meaningful examples from the field of Genetics, establishes that what she calls the 'units of understanding' of a specialised language are constantly evolving and that therefore, consideration of chronology is essential to understand these units. In particular, she shows how scientists effectively use what they know from the past to understand and name the new concepts that emerge in their fields of expertise.

Two other theoretical currents have contributed to enrich and clarify considerably the research work on diachronic terminology. The first theoretical current is based on the recognition of terminological variation, which has been a topic of particular interest since the end of the 1990s, as reflected in Daille et al. (1996), Freixa (2006); Bowker and Hawkins (2006); Desmet (2001) and Pecman (2014), among others. Since 2000, one of the main virtues of this approach is to show that specialised terms are not fixed, but rather should be regarded as variable entities that depend on the field, communication context, situation, and users. This observation of the variability of terms and lexicons has greatly enriched the diachronic dimension of Terminology because it has made it possible to integrate the diachronic variation in a more global approach to specialised languages. Changes that may occur in terms over time are indeed just as much a part of their variability as the context, field or situation in which they are used.

The second theoretical current, based on Textual Terminology (see Condamines and Picton in this volume), has also proved a major contribution to the development of diachronic studies in Terminology, and more precisely, has shown that the study of diachrony in specialised languages can be explored in corpora, using different linguistic clues. The pivotal study by Condamines, Rebeyrolles and Soubeille (2004), which describes lexical variation in the field of Aerospace while taking into account the diachronic perspective, is quite innovative as it is one of the first studies that designs a balanced methodology for the analysis of diachronic corpora. It is also one of the first to highlight the need to combine linguistic and computer requirements in corpus-based diachronic studies.

Within the framework of Textual Terminology, the work of Picton (2009, 2011, 2014a) is particularly innovative in the field of Diachrony, as she is one of the first terminologists to show that changes in a specialised language over short periods of time (even periods of time of less than twenty years) are worth studying. Furthermore, her research in relation with the CNES (Centre National d'Etudes Spatiales, the French space agency) shows that the diachronic perspective can be used as a practical tool to facilitate professional communication. Picton (2009) shows that analysing the changes in aerospace terminology over a very short time period allows senior engineers, who are not always aware that the terms they use evolve over time, to better communicate with young, newly recruited engineers who

sometimes use different terminology. Her work has also contributed to the design of a methodology for the corpus-based exploration of the diachronic perspective, using a set of linguistic clues. Her research has greatly helped to provide specialised diachrony with a detailed and consistent methodology of analysis.

2. The nature of specialised diachrony

This part of the chapter focuses on the nature of specialised diachrony and describes its relevance in exploring the terminologies of specialised fields. Specialised diachrony is intrinsically different from diachrony in general language because it has followed a different path and has served other purposes. When studied in its own right, specialised diachrony brings to light interesting features of specialised languages. For example, it has shed new light on some of the theoretical postulates of Terminology as well as on some linguistic issues that were hitherto neglected by terminologists working on synchrony.

To begin with, specialised diachrony has challenged some of the theoretical postulates of traditional Terminology, which were already questionable from a synchronic perspective but could no longer hold up in the light of diachronic analysis. This is the case of term connotation, for example. Even though most researchers now agree that specialised terms can be connoted, diachronic studies have shown that connotation plays a major role when it comes to the adoption (or non-adoption) of a newly created term.

For instance, Thoiron, Iwaz, and Zaouche (1997) studied the expert use of recommended medical terms in French⁸ over a period of 25 years and showed that some of the terms investigated had not been adopted by specialists because they considered their connotations to be inadequate or inappropriate. This was the case for *nucal* [cervical], which has never been used by experts, who claimed that it did not sound sufficiently scientific. This was also true for the terms *tire veine* [vein stripper] and *eveinage* [vein stripping], which have never been used either, as the experts interviewed for the study said that they sounded “disgusting” (Thoiron, Iwaz and Zaouche 1997, 69).

The study of term connotation over time also shows that an inadequate connotation may give rise to synonymic variation in the lexicon since users often generate other terms to replace the one that they feel are inappropriate. Beeson, Darnovsky, and Lippman (2015) have shown that the term, *commercial surrogacy*, in the field

8. *Recommended terms* refer to the terms that are proposed by the government terminology commissions charged with language planning in France.

of third-party reproduction, had to be replaced by the term *for-profit surrogacy*, because it was perceived as less offensive by parents (2015, 806):

This glossary was amended in a subsequent report eliminating the term ‘commercial surrogacy’, and replacing it with ‘for-profit surrogacy’. [...] A footnote explains that the change was made following feedback from ‘intending parents’ who found the word ‘commercial’ to be offensive.

The study of diachronic variation in Terminology has also challenged the notion of ‘specialised field’. Contrary to the Wüsterian principle that conceives specialised fields as water-tight compartments, diachronic studies on the evolution of terms and concepts have shown that specialised fields (and languages) should be regarded as entities without fixed boundaries, and that a certain amount of conceptual and semantic sharing takes place over time. This was illustrated in Gambier (1987, 3) on acid rain terminology:

Le domaine des dépôts acides fait appel à diverses spécialités (chimie, météorologie, géologie, sylviculture, biologie, mathématiques, droit, étude des matériaux) ; il est à la fois objet de recherche scientifique [...], objet d'interventions juridiques [...] et objet de prises de position publiques d'écologistes, de politiciens, d'industriels.⁹

A number of more recent studies have also described the sharing of concepts and terms across specialised fields that are scientifically very different (Bordet 2013; Fries 2016; Dury 2018, among others). For example, Resche (2013) describes the diachronic evolution of the term *economics* and highlights the disciplinary links that economics has developed over time with fields as varied as Neuroscience and Physics. Previously, Resche (2000, 629) even argued that the theory of Terminology should account for what she calls “interface terminology”, which results from and reflects the increasing interconnection between specialised fields:

By ‘interface terminology’, we refer to a set of terms resulting from the interconnection of disciplines which are not normally perceived as contiguous. [...] As sciences are evolving at a rapid pace, the need for interface terminology will be felt more and more acutely by the specialists and users concerned.

Specialised diachrony has also made it possible to explore new territories in Terminology and to study phenomena hitherto neglected by terminologists. This is the case of term obsolescence, which has been considerably overshadowed by

9. The field of acid rains involves various specialties (chemistry, meteorology, geology, forestry, biology, mathematics, law, study of materials); it is at the same time the subject of scientific research [...], legal interventions [...] and public statements by ecologists, politicians and industrialists (our translation).

the multitude of publications on the creation, formation, and adoption of new terms. The study of neology has always been a major source of interest for terminologists, whereas term obsolescence has only recently aroused the interest of researchers. Dury and Drouin (2011), for example, examined the semi-automatic detection of term disappearance in the field of Terrestrial Ecology over a period of 55 years. Their study proposed and tested a set of linguistic clues, which enable terminologists to explore term obsolescence in corpora. Their results showed that the study of term obsolescence is essential, just as much as neology, in order to update specialised databases.

In the same vein, diachronic variation sheds new light on the question of how experts evaluate terminology and how they comment on it. Picton (2014a) describes how experts frequently express themselves in many different ways about the language they use at work (e.g., by pointing out the inadequacy of a term, by giving details about its meaning, etc.). She also shows that terminologists can use the linguistic comments made by experts on the terms they use, both as an indication that terminology has evolved and as a means to detect new terms resulting from this evolution.

Specialised diachrony has also followed a path of its own since, unlike diachronic studies in general language, it favours studies covering shorter periods of time. Even though some terminology works focus on diachronic studies covering many decades and even centuries (i.e., Van Campenhoudt 1998; Auger 2006; Zanola 2014), the vast majority of diachronic studies in Terminology target short-term changes, taking place in the discourse of the restricted communities of expert speakers. They take into account elements external to language in the analysis of these changes and rely mainly on the study of contemporary states of language, sometimes over very short time periods (less than 10 years).

This growing interest in diachronic studies over short and very short time periods in Terminology has been clearly promoted by the use of specialised corpora and the possibility of analysing them in a semi-automatic way. The use of specialised diachronic corpora in Terminology has radically changed the notion of diachronic variation since it has led to an attention shift from a “historical” to a more contemporary diachrony. It is now methodologically possible to study specialised languages with a precise granularity, over sometimes only a few years, in order to identify, for example, micro-changes, emerging vocabularies, etc. According to Kawaguchi, Minegishi and Viereck (2011, 8):

It is noteworthy that the appearance of large corpora of written texts for some languages has made it possible to analyze quantitatively as well as qualitatively the conditions of historical changes, not only over a long span of time, but also over a short span.

However, the study of diachronic variation in specialised corpora raises a number of research questions that differ significantly from the questions usually linked to the use of synchronic corpora. Are corpora the only means of measuring the diachronic evolution of a specialised language? What methods should be used to detect changes in a corpus? What type of corpus should be used? For all these questions, further research work is still needed, as there are no ready-made answers.

3. The corpus-based study of diachronic variation, a new territory to explore

3.1 The use of dictionaries to explore diachronic variation

Before addressing the question of corpora, the first element that needs to be examined is the use of dictionaries as possible witnesses of the lexical and semantic changes taking place in a specific domain, over time, since many researchers use them to describe and measure the diachronic evolution of specialised lexicons.

Zanola (2014), for instance, analyses a collection of dictionaries and encyclopaedias to describe terminological and conceptual developments in science and technology in the 18th century. She shows that the natural sciences, which developed conceptually, because of rationalisation and experimentation, use terminology that reflects logical observation, induction and deduction, whereas craft techniques, which were in the process of being industrialised in the 18th, use terminology in which standardisation plays a major role. Van der Yeught (2018) also uses lexicographical resources to study what he calls “the process of specialised dictionarisation”, which consists in identifying the emergence of new specialised fields. According to him, when an emerging scientific field becomes conceptually autonomous, the specialised elements that characterise its discourse (lexical units, phraseological elements) are then progressively isolated and identified in lexicographical works, glossaries, and lexicons.

In this case, the use of lexicographical resources to study the changes taking place over the (often long) course of time can be a source of diachronic information on the structuring of specialised fields and the evolution of their terminologies. However, while dictionaries can serve as useful reference tools for identifying changes over long periods of time, the question is whether they can make it possible to observe changes in short or even very short (less than ten years) time periods. Another issue is whether they describe all types of diachronic phenomena in a specialised language, such as changes in the connotation of a term, changes in the co-occurrence of a term, progressive obsolescence of a term or a family of terms, etc.

For these reasons, and because the use of dictionaries alone to study terminological evolution in all its forms is not sufficient, the use of corpora to explore diachronic variation has become very popular among terminologists.

3.2 The pitfalls and limits of exploring diachronic variation in specialised corpora

The semi-automatic exploration of corpora has become an essential exercise for many terminologists in the context of their research on a wide variety of topics (e.g., neology, specialised metaphors, conceptual relations, contexts, etc.). The generalisation of corpus-based studies, however, is also perhaps its main flaw, as the great heterogeneity of corpora compiled, the various types of analysis carried out, and the different methods used make it increasingly difficult to reach a consensus on what makes a reliable corpus analysis. For this reason, what is considered to be scientifically sound corpus-based research by some may not be regarded as such by others. Things become even more complicated when it comes to combining corpus analysis and the diachronic perspective, since many of the criteria regarded as essential elements of Corpus Linguistics in synchrony cannot be duplicated in diachrony. This is the case, more specifically, for the size of the corpus, its representativeness, and the types of analysis performed.

In the very precise context of corpus analysis, diachronic variation still suffers from the predominance of the synchronic perspective since the criteria used for the compilation of terminological corpora and their analysis are essentially criteria which were designed for synchrony. Similarly, as already mentioned, the overwhelming majority of software used for corpus analysis is designed for synchronic studies and only allows for the partial observation of the diachronic evolution of a lexicon.¹⁰

What are the main differing points? First of all, in the case of terminological studies that focus on long and very long periods of time, access to old documents may prove difficult since their format may not be usable for software analysis. Banks (2017) for instance, who studied the evolution of the grammatical structures of scientific articles in the *Journal des Sçavans* and *Philosophical Transactions* between 1665 and 1700, explained that he had to manually explore his corpus of articles since they were only available in image form. Banks (2017, 45) writes:

10. Recent tools developed by specialists in Artificial Intelligence, dedicated to the analysis of lexical embedding, e.g., *Word2vec*, see Mikolov and al. (2013) or *HistWords*, see Hamilton, Jure and Jurafsky (2016), make it possible to map the semantic evolution of terms fairly accurately and seem very promising for diachronic analysis, but they are complex tools to be used for linguists who are not also computer scientists.

In an ideal world, one would choose to analyse the total production of the two journals for the whole of this period. However, such an undertaking implies the availability of the texts in a form suitable for computerized analysis. This is not the case. Although the texts can be found on various Internet sites, they are available only in image form, which renders them unsuitable for computerized linguistic analysis. [...] All these points to the manual analysis as the only solution in the present circumstances.

Depending on the field, the limited access to computer-suitable documents may also apply to short diachrony. For example, most pharmaceutical journals published in English, with one or two exceptions, do not provide access to articles with a publication date prior to 1990. In the same field, in French, it is not possible to access articles prior to 2012.

For all these reasons, and because specialised corpora are by nature circumscribed to specific discourse communities, the representativeness of diachronic corpora in specialised languages may not correspond to the criteria generally mentioned for general language in synchrony. It is therefore important to design criteria that are adapted to the realities of diachronic analysis. For Condamines, Rebeyrolles and Soubeille (2004), the compilation of a diachronic corpus must meet three main criteria or requirements: (i) the requirement of homogeneity (compiling texts belonging to the same textual genre); (ii) the requirement of diachronicity (organising texts over time) and (iii) the requirement of contrastivity (dividing the corpus into several sub-corpora which will then be compared with each other).

From a methodological point of view, another element to be taken into account for the building of diachronic corpora is the temporal delimitation of the corpus, i.e., the choice of a start date and an end date, as well as the division of the corpus into subperiods. To this end, the need to take extralinguistic elements into consideration should be stressed. Indeed, the evolution of specialised lexicons over time cannot be explained solely by taking into account elements that are internal to the language. The role played by historical, social, political, and scientific events are of prime importance. According to Delavigne (2006, 90), "les événements terminologiques sont indissociables des conditions dans lesquelles ils émergent et le contexte sociohistorique n'est pas sans effet sur la constitution et l'évolution du vocabulaire".¹¹ The use of extralinguistic elements to compile a diachronic corpus makes it possible to identify, understand, and explain terminological changes occurring in a specialised lexicon. Kytö, Rudanko and Smitherberg (2000, 86) describe

11. Terminological events are inseparable from the conditions in which they emerge, and the socio-historical context is not without effect on the constitution and evolution of vocabulary (our translation).

the extralinguistic elements used by them in the constitution and the temporal segmentation of the CONCE corpus:

The delimitation of the subperiods was partly governed by the availability of suitable texts [...]. However, the periods also correspond roughly to extralinguistic events of importance in Great-Britain [...]. Period 1 corresponds roughly to the period before the great reforms. Period 2 begins when some reforms (e.g. the First Reform Bill and the penny post) have had time to take its effect, and coincides with a number of other important reforms (e.g. the Repeal of the Stamp Act and the second Reform Bill). Period 3, finally, is more stable, with only a few reforms.

Extralinguistic elements may also help to clarify some of the results obtained in corpus analysis. Delavigne (2006) shows how the discovery of radium by Marie Curie has had a lasting influence on the terminology of Nuclear Physics by introducing a morphological matrix (*radioactivity*, *radioelement*, and *radioprotection*) which is now widely used. Elchacar (2016) shows how the decline in media coverage of alterglobalisation corresponds to a drop in the number of corpus occurrences of the term *anti-globalisation* (2016, 86):

Comme il y a moins d'événements à visibilité médiatique, telles des manifestations aux portes de rencontres économiques, l'emploi des néologismes dans la presse diminue. Ceci expliquerait la baisse de fréquence de « antimondialisation » en particulier, plus lié à la contestation et aux manifestations.¹²

However, using extralinguistic elements to compile and analyse a diachronic corpus is not without pitfalls. First of all, the temporal delimitation of diachronic corpora, as we have seen, also depends on the availability and format of the sources, which sometimes make it impossible to organise the corpus in accordance with the scientific, political, social, or historical context. Secondly, in the case of corpora covering long or very long periods of time, the chronological breakdown of the corpus into historical relevant subperiods can be challenging and can constitute what Monière and Labbé (2010, 805) call a *blind spot* in the analysis. In this case, there is no guarantee that the corpus analyses and the results obtained are not influenced by the purely empirical temporal division of the corpus.

Finally, the corpus-based study of diachronic variation must take into account the fact that the rates at which terminologies evolve may be different, depending on the field, as well as the type of phenomena observed. Thus, for example, if the aim of the diachronic study is to observe how textile terminology has increasingly

12. As there are fewer events with media visibility, such as demonstrations at economic meetings, the use of neologisms in the press decreases. This would explain the decline in the frequency of 'anti-mondialisation' in particular, a term which is more related to protest and demonstrations (our translation).

borrowed medical concepts and terms over time, it is necessary to use a corpus that covers a longer period of time, which would not be the case if the research objective was to study how a certain scientific innovation or event (the sudden appearance of the Covid 19 for instance) has led to changes in the terminology of a specialised field.

4. Conclusion

This chapter aimed at describing the emergence of the diachronic perspective in Terminology and its progressive integration in research works. After a long period of denial, the diachronic perspective, especially since the late 1990s, has aroused the interest of terminologists and has been the subject of a growing, though still relatively modest, number of publications and research studies. These studies have explained theoretical perspectives on the evolution of terms and concepts and have contributed to the development of *ad-hoc* practices and working methods.

They have shown that specialised diachrony is a polymorphic object of study that can be analysed from different points of view, and be based on a wide variety of research objectives. They have also shown that diachronic variation makes it possible to study specialised languages from an original and productive angle. It also provides an opportunity to highlight phenomena which may go unnoticed in synchrony such as the importance of connotation, term obsolescence, rhythms at which different terminological changes occur, importance of extralinguistic reality, etc. All this research work has shown that diachronic variation plays an essential role in the constitution and transmission of specialised knowledge.

However, interest in terminological variation is still recent, and Specialised Diachrony therefore remains a largely underexplored area of research. Many practical questions, particularly those related to the use of corpora to explore diachronic variation, remain unresolved. For example, more research work should focus on designing linguistic methodologies for the analysis of diachronic corpora. Similarly, there are many elements that deserve to be studied from a diachronic perspective, such as term definitions and their evolution, the semantic prosody of terms and families of specialised terms and how they change over the years, the building of terminologies in innovative and emerging fields, etc.

This chapter has thus presented an overview of what has been achieved and what still remains to be done. The glimpse of the exciting new pathways that have not as yet been fully explored will hopefully be an incentive for future research on diachronic variation in Terminology.

Cognitive approaches to the study of term variation

Sabela Fernández-Silva

Pontificia Universidad Católica de Valparaíso

This chapter presents a cognitive approach to research on terminological variation, which posits that term variation is cognitively motivated, since it reflects differences on how specialized concepts are categorized by different individuals in different contexts. Firstly, an overview is given of the theoretical principles that focus on the flexible nature of specialized knowledge categorization and naming. Secondly, we present a classification of the cognitive factors of term variation, which differentiates between systemic and contextual factors. Thirdly, we report the results of a study on the cognitive function of term variation, which indicate that term variation is used as a cognitive strategy to transfer specialized knowledge, but also to avoid repetition, accommodate the text to the audience, or generalize.

Keywords: term variation, intra-textual term variation, cognitive factors of variation, cognitive function of variation, conceptual categorization, specialized knowledge structuring

1. Introduction

Today, terminological variation is a productive and stimulating area of research in terminological studies. All contemporary theories of Terminology agree that variation is an inherent property of terms (Cabré 1999b; Diki-Kidiri 2008; Faber 2012; Gaudin 2003; Temmerman 2000). A growing number of studies focus on the description of the behaviour of variation and its function within the context of specialized communication. Undoubtedly, this situation is the result of the “descriptive shift” that Terminology experienced when it opened its doors to Linguistics. Translators, language planners, lexicographers, and language teachers have become interested in the study of terms in a variety of communicative situations and the development of terminological applications for different types of users and needs.

In such a scenario, variation has proved to be the “touchstone for the understanding of the major issues of terminology research today” (Humbley and Picton 2017, 6), because it is present in virtually all Terminology-related topics. These include the following: (i) how reality is categorized from different subject fields; (ii) how specialized knowledge is structured; (iii) how concepts are named; (iv) how terms are used in different communicative settings; (v) how terms should be represented in applications tailored for different types of users.

Despite its recent history, the variationist study of terminology has gone through different stages. These stages can be related to different lines of research that still co-exist in the landscape of term variation research (Fernández-Silva 2018).

An initial stage, which was influenced by the General Theory of Terminology, posited that variation was an obstacle to effective communication (Wüster 1979). However, at the end of the 1970s, the first reactions against the prescriptivist paradigm appeared in France and in French Quebec, from authors that introduced a sociolinguistic orientation to terminology work, namely Yves Gambier, Louis Guespin, Pierre Auger or Jean Claude Boulanger (Auger 1974; Gaudin 2005). They claimed the need to study the circulation of terms in discourse, taking into account the diversity of social usages (Gaudin 2005, 81). The first variation studies argued against the principle of biunivocity posited by Wüster and showed that term variation existed in different domains and text types and needed to be accounted for in terminology theory and applications. In this stage, the first typologies of term variation were proposed, accompanied by a series of non-systematic reflections about the factors underlying this phenomenon (Auger 1974; Daille et al. 1996; Duquet-Picard 1986; Faulstich 1998).¹ Although today the theoretical status of variation is fully acknowledged, this line of research is still present (Bowker and Hawkins 2006; Depierre 2007; Ferrari 2002; Freixa and Montané 2006). This is due to the fact that negative attitudes towards variation still prevail in certain academic circles (Fuentes-Arderiu 2006; Skopnik-Chicago, Bassaber and Lizana 2020), and has to be counterbalanced with scientific evidence.

This initial stage was followed by a transitional stage, marked by the influence of the sociolinguistic variationist research framework, adapted to the specific characteristics of specialized communication. At the theoretical level, the Socioterminology Theory (Boulanger 1995b; Gambier 1991a; Gaudin 1993; Delavigne and Gaudin in this volume) and the Communicative Theory of Terminology (Cabr  1995, 1999b) are based on the need to account for real term usage and to restore the term’s “right to variation” (Boulanger 1995b, 196), whether in the form of polysemy or synonymy. At the descriptive level, term variation is explored from a variationist model

1. See Freixa (2002 and chapter in this volume) for a comprehensive review of pioneering studies on term variation, which includes an in-depth analysis of the first typologies.

of language, borrowed from general language, which mainly considers dialectal and functional factors.

The adaptation of this model to the restrictions of specialized languages implies focusing on variation parameters that are more prominent in specialized discourse. For example, dialectal variation, or variation among users, pays special attention to socio-professional stratification factors that are specific to disciplinary communities, such as the subject field – or horizontal variation – (Névéol and Ozdowska 2006; Tebé 2005), or the type and level of education (Bertaccini and Matteucci 2005; Picton and Dury 2017). In contrast, other dialectal factors such as space (diatopic variation) or gender have been less studied.

Concerning functional variation, related to differences in the use of terminology in different communicative contexts, there is a consensus that the most influential factor is the degree of specialization of the communication. Consequently, it is one of the best researched in the literature (Delavigne 2017; Depierre 2007; Freixa 2002; Gallardo 2000; Messineo 2002; Tercedor and López 2012). The study of variation in relation to dialectal and functional factors remains a fruitful line of research today, with a growing number of studies that examine different variation phenomena in larger corpora or explore the incidence of a specific factor of variation.

The transitional stage was followed by a consolidation stage, where term variation is studied from a framework that is specific to specialized communication and takes into account the double function of terminology, namely, the representation and transfer of specialized knowledge. It is acknowledged that term variation is not only related to differences across users and usages of terminology, but also to the dynamics of the categorization, structuring, and representation of specialized knowledge.

The cognitive dimension of variation has been highlighted by many authors. The observation of terms in oral and written texts led them to point out the following: (i) the difficulty of establishing clear-cut domains (Gambier 1991a); (ii) the evolution of concepts and conceptual systems over time (Dury 1999, 486; Sager 1990, 13; Temmerman 2000, 230); (iii) the multidimensionality of concept classification (Bowker 1997b, 83; Kageura 1997b, 120); (iv) the multifaceted nature of concepts (Cabré 1995, 2; Zawada and Swanepoel 1994, 268); and (v) the role of language in knowledge construction (Gaudin 1996, 608; Temmerman 2000, 16).

Freixa (2002, 174) reflects on the limitations of the sociolinguistic approach to variation and posits the need to take the cognitive dimension into account. She observes that the distinction between user-related and usage-related variation, proposed by Variationist Linguistics, fails to consider other causes. In her opinion, cognitive causes go beyond dialectal and functional causes and have an important weight in terminology.

These reflections in this consolidation stage laid the groundwork for the conceptual theory of Terminology, which materialized when advances from cognitive sciences regarding the dynamic, experiential, and situated nature of cognition and language were brought to the field by theoretical proposals such as the Sociocognitive Theory of Terminology (Temmerman 2000, this volume) or Frame-based Terminology (Faber 2012, this volume).

This was complemented by other authors whose work centers on the cognitive motivation of term variation. This includes how variation phenomena at the cognitive level can relate to variation at the lexical level (Bowker 1997b; Cabré 2008; Fernández-Silva 2011; Freixa 2002; Tercedor 2004). It also addresses the cognitive consequences of term variation, when term variants highlight different facets of the concept, and as a result, can be used as a cognitive device to build and transfer knowledge in scientific communication (Fernández-Silva 2016, 2019; Pecman 2014; Temmerman 2000). This collective work set the scene for the study of variation from a cognitive perspective, which allows a better assessment of the complexity of this phenomenon by taking into account the three dimensions of the terminological unit, i.e., the linguistic, cognitive, and communicative dimensions (Sager 1990; Cabré 2008).

This chapter presents research on the cognitive approach to term variation. Firstly, in Section 2, an overview is given of the theoretical principles underlying this proposal, which pertain to the dynamic, flexible, and situated nature of specialized knowledge and the multiple motivations underlying the naming of specialized concepts. Secondly, in Section 3, we provide a classification of the cognitive factors of term variation, which distinguishes between systemic factors, related to the variability of concepts and conceptual structures, and contextual factors, related to different temporal, social, and situational scenarios in which specialized knowledge is communicated. Thirdly, in Section 4, we report the results of a study on the cognitive function of term variation, which combines a semantic analysis of intra-textual term variants in Psychology and Geology research articles, and interviews with experts who were prompted to reflect on the functions of term variation. These results revealed that term variation is used as cognitive device that provides information about the concept's features and its relationships with other concepts. However, it is also employed as a communicative strategy to avoid repetition, to accommodate the text to the audience and to generalize text content.

2. The cognitive perspective on term variation: Theoretical foundations

The main theoretical foundations of the cognitive approach to variation are related to the dynamic, flexible, and situated nature of specialized knowledge (Section 2.1), and the multiple motivations in the denomination of concepts and the cognitive functionality of term variation (Section 2.2).

2.1 The dynamics of special knowledge categorization and structuring

Theories and models of cognition, developed for general knowledge by the Cognitive Sciences, were integrated into the field of Terminology (Cabr  2008; Faber 2012; Fern ndez-Silva 2011; L'Homme, Robichaud and Subirats 2014; Temmerman 2000). This transformed assumptions of how specialized concepts are created, structured, understood, and named. The following principles support the study of variation from a cognitive perspective.

2.1.1 *Experientialism*

Experiential realism is a philosophical paradigm that challenges the objectivist epistemological tradition, according to which there is a direct correspondence between conceptual systems and external reality (Gallese and Lakoff 2005, 456).

Experiential realism does not deny the existence of the real world and is “committed to the existence of stable knowledge about the world” (Lakoff 1987, XV). However, it emphasizes the importance of our physical and social experience in understanding the world. Our knowledge of the world is marked by intersubjectivity, because it is mediated not only by the nature of our body, brain, and cognitive system, but also by our experience with the environment and our interactions as social beings. The experiential nature of thought is supported by neuroscientific evidence (Barsalou 2003; Faber et al. 2014; Gallese and Lakoff 2005) and is one of the basic pillars of variation in specialized knowledge and language, primarily because it opens up the possibility that reality can be perceived and conceptualized in different ways, depending on the social and interactional context in which human conceptualizers are embedded.

2.1.2 *Embodiment*

The process of categorizing, structuring, storing, and retrieving knowledge is mediated by our bodily structures and other cognitive capacities, such as perception, attention, memory, sensory-motor activities, emotion, and social interaction (Barsalou 2003; Gallese and Lakoff 2005). Conceptual structures cannot be studied or understood independently from the brain and the body, because the human

bodily apparatus imposes a meaningful structure upon experience and cognition (Lakoff 1987; Zawada and Swanepoel 1994). However, embodiment also has a social dimension, because the experience that is at the basis of embodied concepts not only stems from “our collective biological capacities” but also from “our physical and social experience as beings functioning in our environment” (Lakoff 1987, 267). Therefore, social embodiment emphasizes the role of perspective in the subjective construal of reality (Rohrer 2007, 29) and is at the root of variation phenomena at the cognitive and linguistic levels (Zhang, Speelman and Geeraerts 2015).

2.1.3 *Situated cognition*

According to situated cognition theories (Barsalou 2003; Gallese and Lakoff 2005) “conceptual representations are multi-modal simulations distributed across modality-specific systems” (Barsalou 2003, 513). Therefore, conceptual categories are partial simulations of sensory, motor, and introspective states. As Barsalou et al. (2003, 71) point out, a situated conceptualization of anger is likely to include all the perceptual experiences of interacting with angry people, the actions that the agent could perform to handle the situation, introspective states such as the feelings aroused when the agent experienced anger, and information about the settings where the experience of anger took place.

In that view, conceptual knowledge is not reduced to the storage of decontextualized information about objects or events in amodal representations or arbitrary symbols, detached from experience. On the contrary, knowing a concept is being able to partially reenact it and to simulate previous perceptive, motor, and introspective experiences with the concept. It also entails being able to retrieve the information that best fits the new context, in other words, to “produce a wide variety of situated conceptualizations that support goal achievement in specific contexts” (Barsalou et al. 2003, 71). Situated cognition theories are essential to the understanding of how experience and context shape representation, activation, and acquisition of knowledge structures. In Terminology, they have been used as an important theoretical model to build dynamic specialized knowledge resources (Faber 2011).

2.1.4 *Prototype theory, Cognitive Idealized Models and frames*

Contemporary cognition theories highlight the experiential, embodied, situated, and contextual nature of knowledge. This scenario, marked by diversity, calls for other models of conceptual representation and conceptual structuring that are more coherent with the dynamic and multi-faceted nature of general and specialized knowledge. The traditional model that posited clear-cut concepts and static conceptual structures is not always coherent with the observation of specialized language use in real contexts.

Prototype theory (Rosch 1978; Temmerman 2000 and chapter in this volume; Zawada and Swanepoel 1994), Frame Semantics (Faber 2012 and chapter in this volume; Fillmore 1985), and Idealized Cognitive Models (Lakoff 1987; Temmerman 2000) better represent the internal structure of a concept or the higher-level structures in which concepts are embedded and interrelated, because they acknowledge the importance of culture and context and are more similar to the way knowledge is structured in our minds. A large number of specialized concepts are structured by prototypes, with fuzzy boundaries, whose content is updated depending on the specific situation:

Understanding does not usually happen via essential characteristics and relationships, but via nodes of knowledge, which can have varying levels of complexity, depending on how detailed the understanding is or needs to be in a specific situation. (Temmerman 2000, 36)

Moreover, frame-based models of representation seem to be more adequate, since they reflect that concepts are not only linked by logical or ontological relations, but are typically organized within the context of background actions or events (Barsalou 2003; Tercedor 2011). Therefore, they provide a schematic and general structure that can accommodate other types of relations that are more functional and interactive, such as goal, intended purpose, result of, use, etc. (Faber and León-Araúz 2014, 140).

2.2 The multiple motivations and the cognitive functionality of term variation

This section summarizes the most important theoretical contributions to the cognitive motivation and cognitive functionality of term variation. They are the backbone of the cognitive study of term variation and justify its existence, characterize its specificity and assess its utility in the representation and transfer of specialized knowledge.

2.2.1 *The multiple motivations of term variation*

Various terminological studies have pointed out that motivation plays an important role in term formation (Diki-Kidiri 2008; Kocourek 1991; Sager 1997). In motivated terms, the form of a term reflects part of the concept's content. This means that the "language user is able to deduce, at least partly, the meaning of the term from the analysis of its components" (Sonneveld and Loening 1988, cited in Myking 2009).

In a situation of term variation, it is often the case that a concept is expressed by several (motivated) term variants which reflect different facets of the concept's content and show a particular vision of it. Therefore, term variation can be explained

as a result of a multiple motivation, which takes place in the naming process. These multiple motivations can be analysed from a cognitive perspective. In this sense, they can be related to differences in how language users understand a concept, or how they choose a term variant that reflects a particular view of the concept that is more suitable in a specific communicative context (Bowker 1998; Fernández-Silva 2011; Tercedor 2004). As Bowker (1998, 293) remarks:

An expert may consciously choose to use one term to emphasize a particular dimension at one time, and another term to focus on another dimension at another time, even though both terms are referring to the same concept. Moreover, we submit that this process is both valid and necessary in specialist communication, where there is an emphasis on precision.

2.2.2 *Typology of term variation in specialized discourse*

Another important contribution to the study of term variation is the distinction between variation phenomena at the cognitive level and the linguistic level. This distinction allows the identification of different types of variation and provides a better understanding of how they interrelate and influence each other. Variation can take place at the cognitive level because of different conceptualizations of the same reality by different groups, or by the same group or individual in different contexts. Although this conceptual variation may translate into term variation (see Section 3), this does not always occur (Freixa and Fernández-Silva 2017). In such cases, one can either observe how different senses of a term can be activated in different co-texts (Adelstein 2007), or search for other textual traces of conceptual variation, such as the predications about the concept throughout a text or across texts pertaining to different genres and/or levels of specialization (Ciapuscio 2003; Messineo 2002).

At the linguistic level, there are two types of term variation (Cabr  2008; Fern ndez-Silva 2011):

1. **Variation without cognitive consequences:** When term variants differ formally but not semantically, for example, graphical variants such as *distal estuarine system/DES*, morphosyntactic variants such as *mountain range system/system of mountain ranges*, or lexical variants such as *clay mineral/argillomineral*.² In this case, the information about the concept is the same, and the choice of one or the other has no consequences for receivers and their access to the concept.
2. **Variation with cognitive consequences:** When term variants differ formally and semantically, for example, lexical variants such as *basin filling/*

2. Examples are taken from a study on term variation in Geology and Psychology that is presented in Section 4 (Fern ndez-Silva 2019).

vocanosedimentary filling. In this case, each variant displays different aspects of the concept's content and reflects different viewpoints. This variation has cognitive consequences because it influences the way the receivers access the concept. This second type is especially relevant for the cognitive study of variation because it could be the reflection of variation at the cognitive level. It could also be used as a cognitive device to represent and transfer knowledge in texts (see Sections 2.2.3 and 4).

2.2.3 *The cognitive functionality of term variation*

An important issue for the cognitive study of variation is to defend the cognitive functionality of this phenomenon by showing how it assists the processes of building and transferring specialized knowledge. The fact that different term variants can highlight different aspects of the concept can be used as a way to construct knowledge in texts. In line with this, Temmerman (2000, 227) states that “synonymy and polysemy appear to be functional in the process of progress of understanding”. Polysemy, related to the prototypical structure of some concepts, is an economic resource that increases the “informational density, structural stability and flexible adaptability of the categorial system” (Temmerman 2000, 128). Furthermore, the functional advantage of having term variants that reflect different facets of the concept is that language users can choose one or another depending on which aspect they wish to highlight in a particular context. This, in turn, can contribute to a more nuanced communication. Similarly, Pecman (2014, 1) also argues that scientists “purposefully employ term variation as a means for theorizing and better explaining a given concept”. She claims that rewording is a device for theorizing because it allows users to reconstruct and shape knowledge of scientific concepts within a text. Finally, Fernández-Silva (2016, 2019) has explored the cognitive function of term variation inside texts, whose results are presented in Section 4.

3. Cognitive factors of term variation

Fernández-Silva (2011) developed a classification of the cognitive factors of term variation. This classification (see Table 1) was derived from different studies that analyzed how the dynamicity and flexibility of categorization and conceptual structuring correlated with terminological variation. It distinguishes between systemic factors, related to the nature of concepts and conceptual structures independent of a contextual setting, and contextual factors, which are related to differences in conceptualization among users pertaining to different geographical, temporal or social backgrounds, or to the adaptation of conceptual knowledge to specific communicative contexts and goals.

Table 1. Cognitive factors of term variation

Systemic factors		
	Multidimensionality of the conceptual system	Preliminary causes
	Concept flexibility	
Contextual factors		
Different conceptualizations (User-related variation)	Knowledge evolution	Dialectal causes
	Dialects/cultures	
	Subject fields in interdisciplinary contexts	
	Schools of thought	
	Socio-professional groups	
	Individuals/individual point of view	
Adaptation to specific context (Usage-related variation)	Adaptation to level of expertise of the receiver	Functional causes
	Intention/aim/point of view	

Our claim is that cognitive factors are not an independent block that can be isolated from other variation factors. Instead, they interact with other linguistic, dialectal and functional variation phenomena, and thus should be studied simultaneously.

3.1 Systemic factors

At the systemic level, two properties of conceptual organization are reinterpreted as variation factors. Multidimensionality, a property of conceptual systems where several perspectives of classification co-exist, has been cited as a frequent cause of terminological variation (Bowker 1997b and chapter in this volume; Bowker and Hawkins 2006; León-Araúz and Reimerink 2014; Rogers 2004; Tercedor 2011). Since concept designations often reflect classification perspectives, different viewpoints can lead to different terminological variants.

Moreover, concept flexibility is also a property of embodied, prototypical, and situated concepts, which do not have clear-cut boundaries and cannot be defined by a set of rigid properties. Instead, their definitional characteristics can vary, depending on the situation. Prototypically structured concepts are likely to exhibit term variation. According to Temmerman (2000, 138), “prototypical categories exploit their own polysemic potential”, by adding a new sense by virtue of its resemblance to the prototype. Similarly, Freixa (2006, 64, and chapter in this volume) suggests that conceptual imprecision, another form of conceptual flexibility, is an important cause of denominative variation, which is especially common in novel concepts that exhibit a low degree of denominative and conceptual fixation.

Finally, the two systemic factors included in this classification are a necessary condition for the existence of cognitive variation at the contextual level because

the variability of concepts and conceptual structures explain why different conceptualizations can be possible among different groups of users, individuals or contexts. Therefore, it is necessary to examine the interrelation between systemic and contextual factors.

For example, León Araúz and Reimerink (2014, 660) analyzed how “the multidimensionality of the conceptual system [...] correlates with the adaptation to the level of expertise of the receiver”. They examined which conceptual dimensions, or conceptually-motivated term variants, were preferred in expert communication or semi-expert communication in Psychiatry. Similarly, Dury and Lervad (2008) examined conceptual imprecision in relation to knowledge evolution. In a diachronic corpus of geology texts, they observed that when the concept of ‘petroleum’ first appeared, it was accompanied by a profusion of incidental, often competing synonyms. Picton (2011, 140) also interprets term variation as a clue to the novelty of the concept since co-existing variants suggest that the term/concept has recently appeared in the domain. She claims that term variation is an incidental, temporary phenomenon that will gradually disappear when one variant is finally adopted as the name of the concept.

As suggested in Table 1, a parallel can be drawn between systemic factors of cognitive variation and the preliminary causes of variation in Freixa (2006, 2013, chapter in this volume). Linguistic redundancy and the arbitrariness of the linguistic sign are prerequisites for the existence of variation, because “they are at the origin of variability” (Freixa 2013, 43) and explain “alternative designations stemming from dialectal, functional, stylistic, or interlinguistic causes”. Similarly, multidimensionality and concept flexibility are necessary systemic properties of conceptual organization that allow the existence of other contextual variation phenomena.

3.2 Contextual factors

3.2.1 *Different conceptualizations (user-related variation)*

As mentioned in Section 2.1, situated cognition theories regard conceptual knowledge as experiential, embodied, situated, and therefore, sensitive to the temporal, geographic, social and communicative context where it is created and used. That is why variation in the configuration of concepts and conceptual structures is possible among groups of users or even individuals belonging to different temporal, geographical, and socioprofessional backgrounds. This user-related conceptual variation is likely to have an impact on terminological variation, as several studies have shown. Again, these factors are the cognitive counterpart of sociolinguistic dialectal factors (i.e., diachronic, diatopic, and diastratic), and simultaneously influence the behavior of term variation.

Knowledge evolution has been said to be a common cause of term variation. As Temmerman (2000, 15) states, “it is a fact that concepts evolve over time, as do their designations”. However, Dury (1999, 486) remarks that designations change more rarely, whereas “concepts are constantly changing, turning more complex”. As previously mentioned, various authors have observed that term variation appears more often at the initial stage of development of a discipline or concept. Knowledge is not consolidated yet and the appearance of multiple designations reflect this conceptual instability (Dury and Lervad 2008; Picton 2011).

Term variation due to geographical and cultural differences between language users is a well-recognized variation factor, but in terminological studies it has received little attention. From a cognitive perspective, its effect can be one of the following:

- a. Cultural differences between groups of speakers can motivate the appearance of different conceptually-motivated term variants. This is the case of *mouth ulcer* (Fernández-Silva and Tercedor 2019, 154). It is designated as *afta* [aphta] or *llaga bucal* [mouth ulcer] in peninsular Spanish, whereas in Chilean Spanish, *fuego* [fire] or *calor al estómago* [stomach heat] are also commonly employed, motivated by the popular belief that these ulcers are caused by a sudden increase of temperature in the stomach.
- b. The existence of different geolectal term variants which, in principle, cannot be linked to a different cognitive motivation, can affect the way these concepts are understood because they reflect different aspects of the concept’s content. For example, Schnitzer (2006, 602) reports six term variants with differential geographical distribution in Spanish for the economic term *income statement*: *cuenta de resultados* [outturn account] in Spain, *estado de resultados* [outturn statement] in most Latin American countries, *estado de ganancias y pérdidas* [profit and loss statement] in Peru, *cuenta de ganancias y pérdidas* [profit and loss account] in Argentina, and *estado de utilidades* [income statement] in Panama. As can be seen, these variants reflect different viewpoints on the concept, and might have an effect on how the receiver perceives it.

Socio-professional differences between specialized language users (i.e., diastatic variation) is an important source of variation that correlates with differences in the way concepts are perceived, understood, and structured. Our classification distinguishes four factors that account for the social stratification of specialized domains. Firstly, in interdisciplinary contexts, different subject fields approach the same concepts from different perspectives and with different goals. This has an evident impact on conceptualization and naming (Cabr  1999b; Gaudin 1993; Teb  2005).

For example, Bouveret and Gaudin (1997) described how the term *identity* in Bioinformatics refers to slightly different conceptualizations between biologists and

statisticians. Fernández-Silva (2011) also observed different term choices between biologists and economists regarding fishing terminology. More specifically, terms that are shared by both groups of experts are more neutral and frequent, such as *pescador* and *barco de pesca*. However, each group also exclusively employed term variants that reflect perspectives of conceptualization that were more related to their own domain of specialization, such as *organisms of commercial interest* (biologists) vs. *fishing resources* (economists).

Secondly, the existence of different schools of thought is a frequently mentioned cause of term variation (Bowker and Hawkins 2006; Freixa 2006). Freixa (2006, 65) points out two possible justifications for this. Even though sometimes these differences might be due to alternative ways of perceiving and representing knowledge, in other cases, authors might vary terminology simply to differentiate their school of thought from another.

Thirdly, socio-professional differences between groups of experts working in the same domain have been mentioned as a cause of conceptual and terminological variation (Desmet 2007; Freixa 2006; Gaudin 1993). However, few empirical studies have explored this diastatic factor, and much less from a cognitive perspective. A case in point is the study of Bertaccini and Mateucci (2005), who compared the use of terminology among engineers, technicians and workers employed in a motorcycle factory. According to the authors, terminological choice was related to the socio-cultural status of language users, who “employ the language in a different way on the basis of their professional and personal knowledge” (Bertaccini and Mateucci 2005, 4).

Similarly, Picton and Dury (2017) studied diastatic variation in nuclear medicine and university pedagogy, where different specialists are involved. Although many of their results are not linked to differences in conceptualization, they found a differential cognitive function in the co-existence of terminological variants in the discourse of higher education researchers and academic advisers in the field of University Pedagogy. Denominative variation in advisory websites demonstrates the academic advisers’ intention to “explain the co-existence of these denominations and/or to clearly state the chosen terminology” (Picton and Dury 2017, 72). On the contrary, term variation in journals devoted to higher education research aims to discuss “the development of concepts and denominations and their ability to meet theoretical ends and/or propose new solutions” (Picton and Dury 2017, 73).

Finally, this last type of user-related variation factor concerns differences in categorization among individuals, though not as members of a geographical or socio-professional language community. Differences regarding background knowledge, personal experiences, interests, and goals can yield differences in conceptual configuration and structuring among members of a disciplinary community. This, in turn, can have an impact on term variation. As Bowker (1997b, 277) argues, an

expert in a particular field who wants to write a text using pre-existing terms and concepts might find that the notion that he/she wishes to express may be slightly different from the concepts denoted by the terms that he/she knows. Consequently, instead of using the correct terminological expression for a precise concept, he/she might choose another term reflecting a slight shift in this concept's meaning. This phenomenon is also referred to as "point of view" (Condamines and Rebeyrolle 1997; Gaudin 2003), "vision" (Boisson 2001), or "motivation" (Bowker and Hawkins 2006; Kocourek 1991), among others.

3.2.2 *Adaptation to specific context (Usage-related variation)*

As mentioned in Section 2.1, cognition is not only sensitive to the temporal, spatial, and social context, but also to the situational context where communication takes place. This means that the quantity and quality of conceptual knowledge retrieved by experts from memory depends largely on the parameters of the communicative situation as well as the terminological choices that they make. Therefore, functional variation, related to different usages of terminology, can also have a cognitive motivation.

The most relevant functional factor concerns the cognitive and linguistic adaptation to the level of expertise of the receiver. This has a quantitative impact, since more specialized texts exhibit less conceptual and terminological variation. However, it also has a qualitative one because the conceptual content reflected by term variants can also adapt to the informative needs of the communicative situation. For example, Freixa (2002, 141) observed that in texts with a low degree of specialization, term variation is used as a tool to explain concepts to a non-expert audience.

Furthermore, León-Araúz and Reimerink (2014, 665) showed that term variants of psychiatric diseases selected different dimensions of the concepts depending on the level of expertise of the text genre. For example, the dimensions of symptoms (e.g., *burning-mouth syndrome*) and of discoverer (e.g., *Korsakoff's psychosis*) appeared more frequently in term variants in the expert subcorpus, whereas the body-part dimension (e.g., *Broca's aphasia*) was more frequent in the encyclopaedic corpus that had a lower level of expertise. Finally, Tercedor and López (2012, 248) claim that terminological variation in health-care settings is "a useful tool to bring medical knowledge closer to different audiences", and that "in an interlinguistic and intercultural setting, the awareness of these forms of variation is fundamental in succeeding in providing access to health for all" (Tercedor and López 2012, 264).

The last contextual factor of this classification concerns the conceptually-motivated terminological choices that a language user makes at a given point in discourse. Since term variants can reflect different facets of the concept's content, an expert can choose one or another depending on the facet he or she wishes to

emphasize according to a particular interest or intention. As shown in the next section, the designation of concepts, depending on the facet underlined, can be used to convey information about the concept and its relation to other concepts inside texts.

4. A study on the cognitive and communicative functions of term variation

As previously mentioned in Section 2.2.3, term variation has been attributed a functional role in the process of building and transferring specialized knowledge. The rationale behind this claim is that term variants that highlight different aspects of a concept provide a richer representation of its content and, therefore, can be used as a tool for knowledge representation and transfer. The few studies that empirically researched this hypothesis in Terminology (Fernández-Silva 2013; Pecman 2012b, 2014) drew their observations from corpus analysis. This methodology provides rich descriptive data that reveals to what extent variation is present in texts and the quantity and type of conceptual information that different term variants provide. However, it has limitations when one wants to dig deeper into the motivations and functions of term variation, because it cannot tell us about the process of text production or text comprehension. For example, why do specialists use term variants instead of repeating the same term consistently when writing texts? Are they using term variation as a cognitive strategy to explain the concept, or are they motivated for stylistic reasons, to avoid repetition or for the sake of textual economy? On the other side, does term variation improve or hinder the comprehension of specialized knowledge in texts? Do readers read more easily a text with term variation than a text without it? Those are important questions that need to be addressed when researching on the cognitive function of term variation.

The study presented in this section combined two methodologies in order to study the cognitive and communicative functions of intra-textual term variation in research articles (Fernández-Silva 2016, 2019): (i) the analysis of the incidence and types of term variants in a corpus of 38 research articles in the fields of Psychology and Geology; (ii) semi-structured interviews with six experts, with the purpose of obtaining their views on term variation and its function(s) in the discourse of their disciplines. As shall be seen, the combination of two methods of data recollection helped to better assess the role of term variation as a strategy for knowledge construction and transfer. The semantic analysis of term variants and the opinions of experts who were asked to reflect on the reasons why they used term variation in texts allowed us to confirm that term variation is consciously used as a cognitive strategy and to obtain a detailed classification of the different functions of term variation in texts. The main results are summarized in the following sections.

4.1 Corpus analysis of term variants in research articles

The analysis was conducted on a small corpus of 38 Psychology and Geology research articles (358,425 words) in Spanish. Term variants of the main concepts of each article were detected manually and semi-automatically (Fernández-Silva 2019, 629) and validated by field experts. Then, they were classified according to the type and degree of semantic change from the base term. This involved comparing the conceptual information displayed by terms and variants. Eight variation types were identified (Fernández-Silva 2016) and grouped into three categories of semantic distance. Different cognitive and communicative functions were subsequently attributed to each variation type based on the quantity and type of semantic information provided by term variants. The rationale behind this analysis was that the greater the semantic distance between the term variant and the base term, the greater the probability that term variation performs a cognitive function. Table 2 shows the classification, the frequency distribution in the Psychology and Geology subcorpora (expressed in percentage of occurrences), as well as an example.

Table 2. Classification of term variation according to semantic distance

Semantic distance	Variation type	Geology (%)	Psychology (%)	Example (<i>base terms; term variants</i>)
Maximum semantic distance	change of conceptual configuration	3.2	9.0	<i>paternidad tradicional/modelo hegemónico de paternidad</i> [traditional paternity/hegemonic fatherhood model]
	hypernymic/hyponymic selection	8.7	19.0	<i>afectividad/interacción emocional</i> [affectivity/emotional interaction]
	change of category	11.3	5.6	<i>lago somero/sistema lacustre somero/cuerpo lacustre somero</i> [shallow lake/shallow lake system/shallow lake body]
Medium semantic distance	change of defining feature	3.2	2.1	<i>grupo clínico/grupo TDAH</i> [clinical group/ADHD group]
	change of non-defining feature	3.3	3.1	<i>huella de terópodo/huella de dinosaurio terópodo</i> [Theropod (dinosaur) footprint]
	Reduction	15.9	14.1	<i>nanofósiles calcáreos / nanofósiles</i> [calcareous nannofossils/nannofossils]
Minimum semantic distance	Synonymy	3.1	4.3	<i>mineral arcilloso/argilomineral</i> [clay mineral]
	formal variation	4.1	6.1	<i>sistema estuárico distal/DES</i> [distal estuarine system]
–	Repetition of base term	47.2	36.7	

- **Variants with minimum semantic distance** were found to have a small impact in both corpora. Synonyms account for 3.1% in Geology and 4.1% in Psychology, and formal variants account for 4.1% and 6.1% in Geology and Psychology respectively. The function of these variants is merely stylistic because they do not provide additional information to the reader about the concept. For example, when a multiword term, like *sistema estuárico distal* [distal estuarine system], is replaced by its acronym *DES*, it avoids repetition and favors verbal economy. Other **formal variants** involving morphological changes (e.g., *vínculo inseguro/vinculación insegura* [insecure attachment]) or morphosyntactic changes (e.g., *conjunto de sierras/conjunto serrano* [system of mountain ranges]) can be used to avoid repetition or as a response to the author's expressive needs or carelessness in the consistent use of terminology. Similarly, **synonyms** do not perform any cognitive function because the conceptual content that is underlined is the same. In our corpus they mostly respond to the co-existence of different formation processes (e.g., *mineral arcilloso/argilomineral* [clay mineral]).
- **Variants with medium semantic distance**, which involve lexical changes in the term's modifier, have cognitive consequences that affect the way the concept is represented. The modifier reflects the concept's defining feature(s), so the modification of its constituent elements involves adding, deleting, or changing information about the concept's content.
 - **Reduction**, the most common type of variation in Geology and Psychology research articles (15.9% and 14.1% respectively), facilitates verbal economy and text cohesion. Nevertheless, reductions also provide intercategorical information (Temmerman 2000) about the concept's position within the conceptual structure, since they are, conceptually speaking, hypernyms (*nanofósiles calcáreos/nanofósiles* [(calcareous) nanofossils]).
 - **Term variants reflecting a different defining feature** (*grupo clínico/grupo TDAH* [clinical group/ADHD group]) or **selecting an additional, non-defining feature** (*huella de terópodo/huella de dinosaurio terópodo* [theropod (dinosaur) footprint]) provide information about the intracategorical aspects of the concept – its internal characteristics – which are useful for explaining its content (Temmerman 2000). However, their frequency is low in both corpora (6.5% in Geology and 5.2% in Psychology).
- **Variants with maximum semantic distance** are the most frequent category in both corpora (23.21% in Geology and 33.67% in Psychology).
 - **Variants selecting a different category** in the head element of the term provide different, though complementary, perspectives of categorization of a given concept. For example, the variants, *lago somero* [shallow lake], *sistema lacustre somero* [shallow lake system], and *cuerpo lacustre somero* [shallow lake body] reflect three perspectives of the conceptualization of

a lake – as a land-form, as a complex ecological system, and as a body of water –. Combined within a text, they provide a richer representation of its content and allow readers to better understand it.

- **Hypernymic and hyponymic variants** (8.7% in Geology and 19.0% in Psychology) have a cognitive function because they specify the intercategorical relationships of concept with others within the domain. For example, when the term *afectividad* [affectivity], is replaced by its hypernym *interacción emocional* [emotional interaction], it provides information about the classification of this concept in Psychology.
- **Variants with a change in the conceptual configuration** involve a change in the term's head and modifier. They can perform a cognitive function since they give additional information, as in *paternidad tradicional/modelo hegemónico de paternidad* [traditional paternity/ hegemonic fatherhood model]. However, a rhetorical function was also identified, especially in Psychology research articles where its frequency is considerable (9.0%). Term variation shows a different pattern of behavior in different sections of the article (Fernández-Silva 2016). In the introduction and conclusion, term variation is used to explain the concept and its position in the conceptual structure (for example, *personas con VIH/SIDA/pacientes diagnosticados con VIH/Sida* [patients diagnosed with HIV/AIDS]). In the methods and results section, it links the theoretical concepts explained in the introduction section with the specific sample investigated in the study, which is representative of the general concept. For example, *HIV patients* is replaced by *sujetos del estudio* [subjects of the study] or *participantes del estudio* [participants], which involve a change in the conceptual configuration of the term.

As can be seen, corpus analysis provided useful information regarding the incidence of term variation within texts as well as the type of conceptual information provided by different term variants. From this, it can be inferred that some variation types are more likely to perform a stylistic function, because they do not provide new semantic information, whereas others are more likely to perform a cognitive function, because they inform about different facets and relationships between concepts. However, important questions remain as to whether the writers of texts are aware of these communicative and cognitive functions and make use of term variation in a purposeful way. This is especially relevant for the cognitive function, because Freixa (2005) in a previous study that also explored the functions of term variation by interviewing experts, stated that they were only conscious of the need to vary for stylistic reasons, mainly to avoid repetition or for the sake of verbal economy. The second part of our study focused on these questions.

4.2 Perceived functions of term variation according to experts
(interview analysis)

Oral interviews were conducted with three experts from each field. Firstly, they were asked to read two excerpts from a research article and reflect on the terminological choices and the reasons why specific term variants were chosen. Secondly, they read two excerpts from one of their own research articles, where they had used different term variants for a concept, and explained their choices. Thirdly, they were asked to read 12 statements expressing different opinions on term variation and to mark those statements they agreed with.

Table 3 shows the six functions that were identified by experts during the interviews, which were grouped into the broad categories of cognitive and communicative functions. Three communicative functions were identified in the interviews. The most common was to avoid repetition, which was elicited 50.94% of the times by psychologists and 35.14% by geologists. For instance, one psychologist affirmed that “it is important to be careful when one writes to avoid repetition. I hate repeating myself”. The role of variation as a device for adapting the text to the intended audience was also identified, although psychologists (7.55%) were more aware of this function than geologists (2.70%). It was expressed in statements such as “I guess that the author used all these terms to bring the text closer to readers”, pronounced by a psychologist. Finally, the rhetorical function, which substitutes a general concept for a more specific sample that is representative of the concept in the study, was also mentioned by the experts, though only twice by each group. For example, a psychologist explained that he/she replaced *4th to 7th grade schoolchildren* by *participants* “to situate them as participants of the study”.

Table 3. Cognitive and communicative functions of term variation according to experts

		PSYCH (%)	GEOL (%)
Communicative	Avoid repetition (stylistic)	27 (50.94)	13 (35.14)
	Adapt text to intended audience (functional)	4 (7.55)	1 (2.70)
	Generalize and/or particularize (rhetorical)	2 (3.77)	2 (5.41)
Cognitive	Provide intracategorical information	9 (16.98)	11 (29.73)
	Provide intercategorical information	4 (7.55)	4 (10.81)
	Favor conceptual clarity	7 (13.21)	6 (16.22)
		53 (100.00)	37 (100.00)

The other three functions identified by the experts were labelled as cognitive, and they involved using terms to better explain the concepts, and improve the representation of specialized knowledge in the text. In the first place, term variation is

used to provide intracategorical information about the features of a given concept. This function was the second most common in Geology (29.73%), and Psychology (16.98%) after avoiding repetition. For example, one geologist explained that she/he employed the term variant *paleo-volcanic edifice* instead of *paleo-volcano* “so the reader could picture the structure of the volcano”. Term variation was also used as a device for providing intercategory information, or about the relationships a concept bears with other concepts within the conceptual structure of the domain. It was recalled 7.55% of the times by psychologists and 10.81% by geologists. For instance, a geologist declared that she/he replaced *debris avalanche* by *flow* and *avalanche flow* because “they are a type of flow”, and because “it is a dynamic process and I wanted to convey the idea that it is a process”. Finally, different statements where experts considered variation as a means for better explaining themselves were labelled under the third general cognitive function of “favoring conceptual clarity”. This was mentioned 13.21% of the times by psychologists and 16.22% by geologists.

In conclusion, the interviews confirmed the function of term variation as a device for knowledge representation and transfer in texts. Experts from both fields could recall a variety of communicative and cognitive functions when they analyzed term variation in real texts. This supports the hypothesis that term variation is an important cognitive and communicative device, and points to the need to study term variation from a cognitive-functional perspective, taking into account the cognitive, social and communicative factors involved in specialized language use.

5. Conclusion: Challenges for a cognitive study on term variation

This chapter has presented the main principles of the cognitive approach to term variation. After a brief history of the study of term variation, the theoretical principles of a variationist theory of terminology were discussed. These principles highlight the dynamic and situated nature of specialized knowledge as well as the typology, behavior and functions of term variation in specialized communication.

Cognitive factors of variation include systemic and contextual factors related to different conceptualizations among users and usages of terminology. These factors should not be studied in isolation, but rather in correlation with other dialectal and functional factors. Finally, a study was reported, which explored the cognitive and communicative functions of term variation in specialized texts. The results showed that term variation is not only used as a communicative device to avoid repetition and adapt the communication to the intended audience, but also to provide information about the concept's content and its relationships with other concepts. In our opinion, this review shows that term variation is a phenomenon with

linguistic, cognitive and communicative origins and implications. Furthermore, it is a well-established research area in terminological studies, supported by strong theoretical principles.

Future directions and challenges for the cognitive-based research on term variation include the following:

- More corpus-based, empirical studies are needed that explore the behavior and functions of term variation in relation to cognitive and socio-communicative factors. Comparative descriptive studies between disciplines would explore not only formal differences between term variants but also different causes and functions of term variation according to the discipline.
- Studies focusing the incidence of specific factors that have been under-researched should be encouraged. Variation factors that are in need of more empirical evidence include social factors (e.g., gender, geolect or socio-professional group) and cognitive factors (e.g., school of thought or individual point of view).
- Corpus-based studies on term variation need to be complemented with data collected with other methodologies. Although observation of term variation in specialized corpora allows researchers to characterize term behavior in real language use contexts, it has limitations when it is necessary to explore the causes and effects of variation. For example, other sociolinguistic methodologies such as interviews or attitudinal questionnaires can be used to elicit information and query users of terminology about their motivations, attitudes, and perceived functions of term variation. Furthermore, experimental studies that investigate production, processing or reception, could help to better assess the role of term variation in specialized knowledge construction, understanding and acquisition.

These are only some of the numerous lines of research that are in need of a more in-depth study in order to understand the phenomenon of term variation in all its complexity.

Terminological growth

Kyo Kageura

The University of Tokyo

The generation of new terms involves various factors at different levels and can be described and modelled from different points of view. This chapter analyses approaches that model terminological growth by describing the creation of new terms in relation to the terminology of a domain as a set or as a system, and/or in relation to the conceptual system that it represents. We first examine the concept of terminological growth in contrast to the description and modelling of term formation, and then describe the approaches used to model terminological growth. Finally, we discuss implications for practical issues in terminology processing.

Keywords: terminological growth, term formation, terminological sphere, terminology network, realistic possibility of existence of new terms

1. Introduction

With the constant growth of knowledge in many research domains, new terms are constantly created and some terms become obsolete. While terms need to be as stable, consistent and systematic as possible (Felber 1984; ISO 2000), they are at the same time dynamic and flexible. Some terms go through change in the concepts they represent. We encounter variations of terms in texts.

In this chapter, we deal with terminological growth, which is concerned with how terminology grows. An essential point here is that vocabulary and thus terminology consists of a set of terms and constitutes an independent sphere in itself. This chapter introduces work that tries to characterise its growth. This contrasts with approaches that deal with dynamics of terms taking into account wider factors such as usage in texts or social context (Temmerman and Campenhoudt (2011) contain some of the important articles belonging to these approaches), with those that deal with neologies (Cabr , Estop  and Vergas-Sierra (2012) cover a range of articles that address neologies), or with automatic term processing that generally claims to target new terms or term variations (Drouin et al. 2018).

This chapter is organised as follows. In Section 2, we clarify the position of the study in terminological growth thus defined within the studies of terms and terminologies. In Section 3, we elaborate on the target of the study in terminological growth, taking the study of term formation as a point of departure. Section 4 is devoted to descriptions of concrete approaches developed so far to describe and model terminological growth. Section 5 is devoted to issues and implications. Section 6 concludes this chapter.

2. The scope and position of research in terminological growth

Let us start from clarifying the sphere of Terminology and how it is addressed in the study of terminology as a prerequisite for consolidating the concept of ‘terminological growth’ and for making its theoretical status clear.

Sager (1990, 58), a well-established textbook of Terminology, introduces three different dimensions for studying terminology, i.e., cognitive, linguistic and communicative, and states in relation to the communicative dimension as follows:

[B]y being studied in the context of communicative situations, terms are no longer seen as separate items in dictionaries or part of a semi-artificial language deliberately devoid of any of the functions of other functional items. The increasing tendency to analyse terminology in its communicative, i.e., linguistic context, leads to a number of new theoretical assumptions and also to new methods of compilation and representation.

Since the 1990s, we have witnessed important extensions of terminological research in this direction, e.g., capturing terms in texts, terms in discoursal context and terms in relation to their social functions (e.g., Pearson 1998; Temmerman 2000). Dynamics and variations of terms have also become an important topic of research within this trend (e.g., Temmerman and Campenhoudt 2011; Daille 2017).

Observing the table of contents of this monograph reveals an interesting point in relation to the status of ‘terminology,’ i.e., “the vocabulary of a subject field” (De Bessé, Nkwenti-Azeh and Sager 1997). On the one hand, the first part is devoted to the General Theory of Terminology (GTT), which puts emphasis on concepts and conceptual systems and thus on terminology as a set or system that represents the conceptual system, rather than the use of terms in texts. A part on Knowledge-based Terminology follows the GTT part. On the other hand, most of the following parts covers study of terms and terminologies from a wider perspective, which reflects the state-of-the-art of terminological research. Another interesting point is that such modifiers as textual, socio-, or cultural are explicitly used, when terms and/or terminologies are studied in relation to wider contextual factors. These expressions

indicate that terminology as the vocabulary of a subject field is always and already assumed as a baseline in talking about terminology and terms. We have witnessed the development of the study of terms in relation to texts, society and culture. Most of them do not target 'terminology' *per se*.

We are also witnessing the refinements in analysing and describing terms and concepts by using theories and methods introduced in related fields such as Linguistics and in ontologies (Faber 2015; L'Homme 2020a). These contributions lead to new ways of constructing and managing term bases (e.g., DicoEnviro; DicoInfo; EcoLexicon). In a sense, they can be regarded as an extension of traditional Terminology typically represented by GTT. In these studies, 'terminology' itself is not the theoretical target itself, either. They develop theories to analyse and describe terms and define conceptual or terminological structures (cf. L'Homme 2020a), and the sphere of 'terminology' manifests itself through constructed term bases. The same is true for more ontology-oriented approach to terminology (e.g., Roche 2012c).

The concept of 'terminology' is in general not addressed as a reflective target of theoretical study, though terminologies, i.e., concrete manifestations of 'terminology,' are the main target of practical terminology work and sometimes important target of applications of theoretical research. The question inevitably arises here: is it not that the concept of 'terminology' is not directly addressed because it is not of primary importance? After all, terms occur and function in texts, and terminologies are compiled based on these terms that actually occur in texts. Kageura (2012) argues that, while empirically terms in texts precede a terminology as a set, the concept of 'terminology' is required for the concept of 'term' to be established, which is essential to identify actual lexical items as terms. This framework is depicted in Figure 1.

There seem to be two reasons that 'terminology' tends not to be explicitly set as a target of study. First, the concept of 'terminology' remains abstract unless it is manifested as actual terminologies. These terminologies are constructed through laborious human endeavour, and thus tend to be regarded as secondary products compared to terms in texts and they themselves tend not to be regarded as target of theoretical research (Wilks, Slator and Guthrie 1996). Indeed, constructing terminologies is one of the main tasks for practical terminology work but not for terminology research. A number of international or regional standards or guidelines exist (Chiocchetti and Ralli 2013; Euro TermBank Consortium 2006; Galinski and Goebel 1996; ISO 2000, 2001, 2007, 2009, 2010), which aim at regulating, coordinating or standardising the process of constructing terminologies and incorporating terms into terminology databases.

Figure 1 also implies two important points in relation to the status of 'terminology' and terminologies. First, individual lexical items used in texts that represent

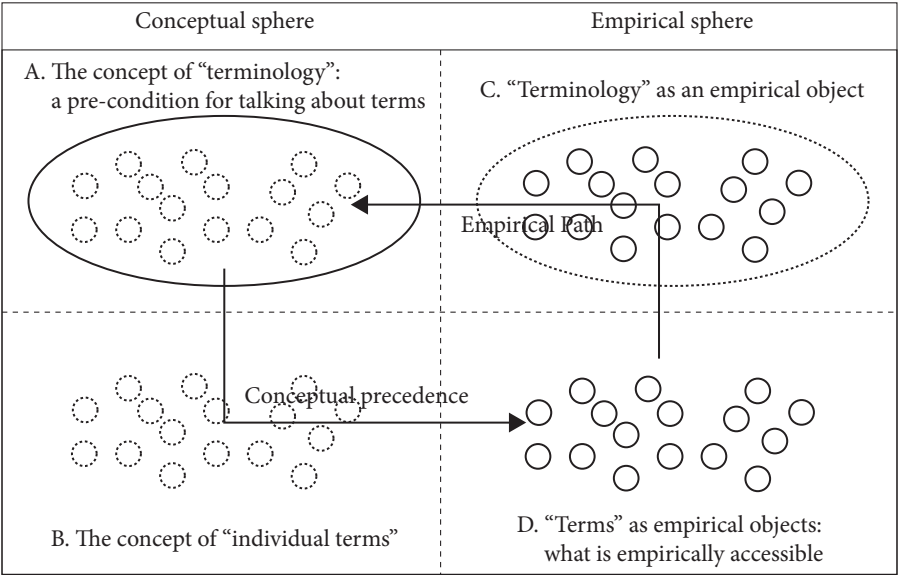


Figure 1. The position of terminology and terms, from Kageura (2012, Figure 1.3)

specific concepts in the discourse of the domain are not necessarily terms of the domain. For those lexical items to become terms, they have to be shared, i.e., collectively approved and accepted, in the domain. This fact, together with the theoretical precedence of the concept of ‘terminology’ to actual terms in texts, in turn shows that a terminology of a domain compiled as an empirical product is autonomous from texts of the domain, even if entries are taken from the texts.

We can now depict the position of the sphere of Terminology within which terminological growth is defined more concretely. Figure 2 shows this position of terminological growth within a simplified relationship between technical discourse

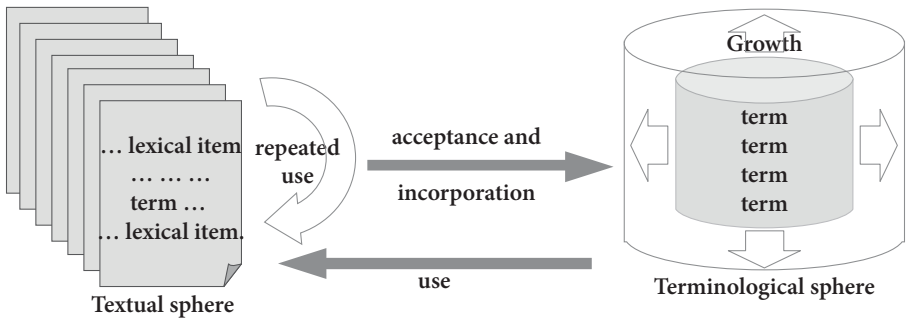


Figure 2. The status of terminological growth in terminological sphere and textual sphere

and terminology. The figure indicates that it is at the same time abstract and concrete. Abstract, because the concept of ‘terminological growth’ can be established only through theoretical abstraction in that it captures the incorporation of new terms in terminology from the point of view of the terminological sphere, although to what extent one takes into account social and discoursal dynamics in the process is a matter of theoretical and/or practical choice. Concrete, because terms are out there as a reality of the domain and the number of terms within a terminology increases in reality.

We do not take into account social and/or textual factors in terminological growth in this chapter. Related aspects are covered, though not directly in relation to terminological growth, in other chapters. As a corollary, interactions between ordinary language and technical terms are not covered either (cf. Budin and Oeser 1995; Meyer and Mackintosh 2000). In other words, we do not address the study of the evolution of terminology in general (cf. Rey 1995) but focus specifically on growth of terminology as seen within the terminological sphere. What is addressed, then, is potential of growth that terminology embodies in its structure at a given synchronic state.

To carry out the study in terminological growth in concrete terms, we need to start from actual terminologies. These are reflectively constructed products created by terminologists, subject specialists, lexicographers and/or sometimes researchers, and may only cover part of terminologies. These are thus the surrogate of ‘terminology’ of a domain, which is the theoretical target of the study in terminological growth. This is an inevitable theoretical and practical condition of the study in terminological growth as defined here.

3. Term formation and terminological growth: Problem setup

This section elaborates on the concept of ‘terminological growth’ and clarifies the core question corresponding to research in terminological growth. We use studies of term formation as a point of departure and proceed to clarify research in terminological growth.

3.1 Term formation

For a terminology to grow, new terms should be created. How individual terms are created is addressed in the study of term formation (e.g., Cabré, Estopà and Vargas-Sierra 2012; Humbley 2018a; Sager 1990, 1997). Here, after we briefly review basic term formation mechanisms/patterns, we summarise how complex term formation patterns are described. Through these, we postulate the research question that studies of term formation tries to answer.

Sager (1990, 1997) distinguishes two types of term formation, i.e., primary term formation and secondary term formation. In primary term formation, a new concept is named with no direct linguistic precedent, although there may be motivation in formation regulated by the concept itself and the concepts surrounding it. It is basically carried out in a single language, although there may be exceptions in which new concepts are named in the context of establishing, for instance, new treaties in official languages of the United Nations in parallel (Temmerman 2013). Secondary term formation, on the other hand, “occurs when a new term is created for a known concept ... as a result of knowledge transfer to another linguistic community” (Sager 1990, 80). As English has been used as a *de facto* lingua franca in many subject domains, primary term formation is expected to occur more frequently in English, whereas secondary term formation is common in other languages, including borrowings.

Given a new concept or a new term in different languages, the process of forming terms and the driving force behind the establishment of term forms may vary. Classical Terminology Theory emphasises the systematicity of terms in relation to the conceptual system that they represent (Felber 1984; Picht and Draskau 1985), while recent research has shown that different factors can play an important role in forming terms. Humbley (2006), for instance, argues that in a certain setup in secondary term formation, metaphor can play an important role. Linguistically, derivations (e.g., *decontamination*, *termhood*), abbreviations (e.g., *COVID-19*, *RT-PCR*), compounding (e.g., *database*, *keywords*), and composition or complex term formation (e.g., *information retrieval*, *directed acyclic graph*), and borrowings (e.g., English *typhoon* is from Japanese 台風, and Japanese データ is from English *data*) are some of the important mechanisms observed in forming terms (Sager 1990, 1997; Valeontis and Mantzari 2006).

From a more regulatory point of view, ISO (2000) lists desirable criteria for term formation as principles of term formation: transparency (e.g., *thermal noise* rather than *Johnson noise*), consistency (e.g., *nylon*, *orlon*, *dacron*, *rayon*, i.e., end in *on* for synthetic fabrics), appropriateness (e.g., *nuclear energy* rather than *atomic energy*) and linguistic economy (e.g., *term bank* rather than *terminological data bank*).

In most languages, new terms tend to be formed in such a way that the created terms reflect the positions of the concepts they represent in the conceptual system of the domain (though this level of description is not necessarily incompatible with the fact that such factors as metaphor, for instance, are involved). Reflecting this tendency, in many languages, it is observed that 70 percent to 80 percent of the terms in terminological resources of many domains are complex (Cerbah 2000; Justeson and Katz 1995; Kageura et al. 1999; Nomura and Ishii 1988). They are thus relatively motivated (Saussure 1916), as shown in the following examples (Kageura 2015):

translation
human translation
machine translation
statistical machine translation
rule-based machine translation
machine translation by analogy principle

The forms of these terms reasonably reflect the concepts they represent, i.e. the heads or the nuclei represent the genus and the modifiers or the determinants represent specific characteristics. Thus we can approximate the conceptual hierarchy by resorting to the form of terms, as shown in Figure 3. A new term, *neural machine translation*, can also be judged to belong to the third level of the hierarchy in Figure 3, by looking at its form.

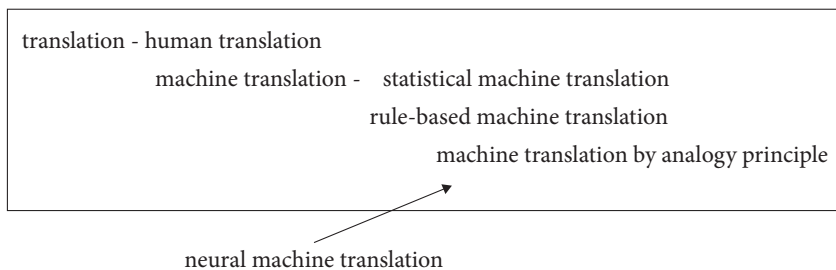


Figure 3. Relatively motivated set of terms

Note that the systematicity of complex term formations vis-à-vis concepts they represent is not complete or rigid. For one thing, complex term formation is not the only method of term formation. For another, complex terms, even if they follow systematic patterns, do not necessarily reflect conceptual system rigidly. For instance, *statistical machine translation* and *neural machine translation* are closer to each other and together contrast with *rule-based machine translation* and *machine translation by analogy principle*. Also, translation carried out by humans and what is understood to be translation in machine translation research are different (Kageura 2019), thus *human translation* and *machine translation* are not indeed precise siblings under *translation*. Terminology exists in the tension between tendencies towards systematicity and towards flexibility, and term formation reflects this (Rey 1995).

As complex term formation is the dominant pattern of term formation, much work has been carried out to capture patterns of complex term formation. Linguistic patterns of terms have been a target of research and description both in themselves (e.g., Sager 1990; Valeontis and Mantzari 2006) and in the context of such applications as automatic term extraction from texts (e.g., Ananiadou 1994; Justeson and

Katz 1995; Daille 2017; Jacquemin 2001; Nakagawa and Mori 2003). Patterns of complex term formation are also described in relation to the conceptual categories of terms and of their constituent elements (Pugh 1984; Kageura 1997a, 2002). Two essential points of observation in these studies are:

- Characteristics of constituent elements used to make a term;
- Characteristics of combination or modifications used to integrate the constituent elements to make a term.

As shall be seen, these are the basic factors in modelling terminological growth as well.

The work of describing patterns of term formation, irrespective of whether they are linguistic or conceptual, contributes to the clarification of the range of possible forms that terms can take. Put differently, these descriptions try to answer the following question:

(Question I) What form do terms take when they are created?

This is a valid and important question, but both theoretically and practically, an essential trait of terminology is left unaddressed, i.e., a terminology as a set consists of concrete terms and does not consist of forms of possible terms. This point is theoretically important because, as previously mentioned, the concept of terminology precedes the concept of terms. Indeed, when terminologists work on creating and managing terminological databases, they deal with concrete terms and not term forms. Computational applications such as automatic term extraction rely on the definition of term forms, but their goal is to obtain a set of new concrete terms. For that purpose, they use textual corpora to substantiate patterns of term formation.

3.2 From term formation to terminological growth

Research in term formation does not address the area of terminological growth depicted in Figure 2, as the former is essentially concerned with allowable forms of terms. To shift the focus of concern from possible forms of terms to possible terms, we can postulate the research question as follows:

(Question II) What range of terms is expected to be created in a given terminology?

This question corresponds to the area of terminological growth as depicted in Figure 2. This is an essential theoretical shift, somewhat analogous to what Foucault (1968, 17) stated, when he contrasted his analysis of discourse with Linguistics:

La question que pose l'analyse de la langue, à propos d'un fait de discours quelconque, est toujours: selon quelles règles tel énoncé a-t-il été construit, et par conséquent selon quelles règles d'autres énoncés semblables pourraient-ils être construits? La description du discours pose une toute autre question : comment se fait-il que tel énoncé soit apparu et nul autre à sa place?¹

Foucault's description of the discourse analysis seems static, but Foucault (1969) explicitly observed that statements are at the same time concrete yet not fixed. Though statements and terminologies are different constructs both in their substance and in their theoretical status, this analogy is helpful.

In any case, to the extent that the terminological sphere is identified as constituting an independent sphere, its growth makes it an essential theoretical target of research. The work in term formation, important as it is, does not address terminological growth. It addresses Question (I) in 3.1, not Question (II) in this section. If the description of term formation patterns is pursued to a detailed degree so as to address Question (II), the end result will be listing all the combinations of linguistic items in the terminological data used in the study, which effectively make the study as a natural history of terms (Kageura 2002). Indeed, such work has been and continues to be carried out by terminologists, i.e. collecting terms and compiling and managing terminologies. Though this work is essential, how terminologies grow is outside its scope, with some exceptional cases in which language planners including terminologists take part in the naming of concepts as in some policy planning in the European Union (EU).

Unlike studies of term formation, research in terminological growth assumes terminologies as a set, consisting of actual terms rather than of their forms. Modelling its growth, therefore, needs to target the range of the "realistic possibility of existence" of terms within terminology (Kageura 2002). This contrasts with the possible forms of terms.

The realistic range of terms that are expected to be created, in one understanding, would reflect research trends in the target subject domain. This depends not only on the research in the subject domain in its Platonic sense but on social trends. As a terminology of a domain represents its conceptual system, its reality may depend on research trends but its potential is not expected to do so. For instance, in the field of Natural Language Processing, we are witnessing rapid growth

1. The question that the analysis of language raises, in the face of a certain fact of discourse, is always: from what kind of rules was this statement constructed, and, consequently, from what kind of rules can other statements, that resemble this one, be constructed? The description of discourse raises a completely different question: how is it that this statement, and nothing else in its place, appeared? (my translation).

in research that applies deep learning methods, and many new terms have been introduced, such as *LSTM*, *transformer*, *attention mechanisms*, etc., while the speed of development seems much slower in logical or symbolic approach to language processing. This does not necessarily mean that the latter approaches do not have the potential of creating new concepts; it may just be reflecting the fact that this topic is less trendy.

Here, we do not directly take into account factors related to research trends, even though they are important for language practitioners including terminologists and translators, and also for such applications as automatic term extraction, trend detection (e.g., Hwang et al. 2011), and technological watch (e.g., Katsurai and Ono 2019). It is not because these factors are irrelevant, but because we intend to consolidate the core structural part of terminological growth here. Taking into account research trends naturally leads us to take into account diachronic aspects of creation and growth of terms, and given that the structural approach to terminological growth is still in its embryonic stage as we see below, we safely save the task for future work. We thus focus on structural and synchronic approach to capturing terminological growth.

4. Modelling and describing terminological growth

The core research question about terminological growth has now been formulated as Question (II) in 3.2. It targets the realistic range of possible terms within terminologies, from structural and synchronic points of view. This section examines existing and possible approaches to address this question. Section 4.1 introduces a way to anchor term formation patterns to terminological structure in a qualitative manner. Section 4.2 deals with a quantitative framework and methods of modelling growth patterns of constituent elements of terms in terminologies. Section 4.3 extends the modelling of growth patterns to modifications and conceptual specifications.

4.1 Anchoring concepts to the conceptual system

Structural or synchronic modelling can be divided into qualitative and quantitative approaches.

Addressing the issue from a qualitative perspective, Kageura (2002) proposed to extend the description of term formation patterns to the description of patterns of terminology formation, in which the conceptual nucleus and other conceptual characteristics are anchored to relevant categories within the conceptual system of the domain. This is depicted in Figure 4.

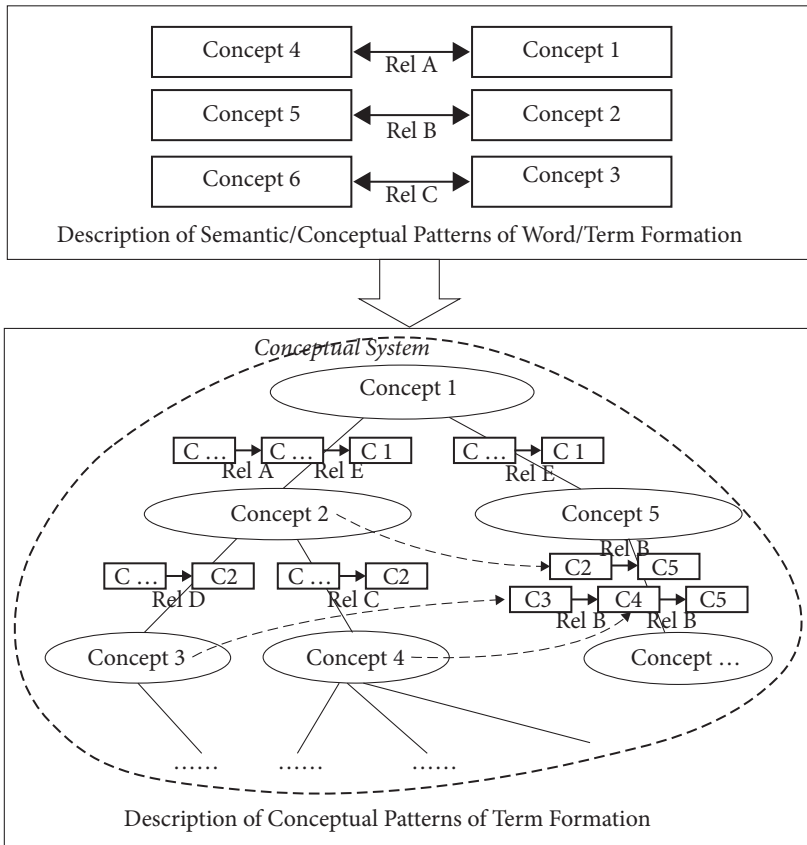


Figure 4. Term formation and terminology formation, from Kageura (2002, Figure 3.1)

In relation to this shift in the point of view, Kageura (2002) introduced two concepts that constitute the keys to move from the study of term formation to the study of terminological growth. The first is the “classificatory patterns” of term formation. Complex terms, which constitute the majority of newly formed terms, are created by narrowing down existing concepts; in the process, how the concepts are further specified depends on other sibling concepts. In Figure 3, the introduction of the form of the term *statistical machine translation* is not just determined by the concept it represents, but is also affected by its sibling terms, i.e. *rule-based machine translation* and *machine translation by analogy principle*. The second is the rearrangement of term formation patterns into “conceptual specification patterns” within the conceptual system in the subject field, based on the term formation within the terminology.

Though these extensions are theoretically a step towards grasping terminological growth, the scheme has an inherent limitation in the granularity of descriptions,

because categories (in this case conceptual categories) introduced to describe classificatory patterns and conceptual specification patterns are necessarily a generalisation from that concepts that individual terms and constituent elements of terms represent. Using conceptual or linguistic categories are theoretically important to capture general patterns of term formation, but they are essentially the means to answer Question (I), i.e. possible form of terms, and not the means to answer Question (II), i.e. realistic range of possible terms.

4.2 Growth patterns of constituent elements of terms in Terminology

When we talk about terminological growth as distinct from term formation patterns, we are concerned with individual concrete terms that actually exist and that can potentially exist. Unlike the description of term formation patterns in which attributes of terms and their constituent elements are analysed, this research is obliged to observe substance, i.e. individual terms and individual constituent elements of terms that represent individual concepts and individual conceptual characteristics. This brings us back to previously mentioned conundrum. More specifically, since terminology researchers (and terminologists) are not in a position to create new terms, what can they do in relation to terminological growth, other than just listing existing terms and adding new terms each time that they occur? Although not explicitly recognised as such, this problem shares the basic structure with one of the problems in Natural Language Processing (NLP), namely, how NLP systems should handle items that do not occur in training data (cf. Manning and Schütze 1999). Although NLP tasks, such as parsing, are concerned with forms, many applications including machine translation and information extraction have to deal with individual items.

A standard framework to deal with this issue is to leave room for the unseen items and estimate how much unseen events are expected and in what way (Baayen 2001; Manning and Schütze 1999). Although subword or character units are used in neural language processing (Faruqui et al. 2017), they do not address the issue of unseen items *per se*. A range of theories and methods have been developed so far in dealing with unseen items from the distributions of existing items (e.g., Baayen 2001; Good 1953; Good and Toulmin 1956; Katz 1987; Ney, Essen and Kneser 1994). As the research in terminological growth is concerned with specifying the range of terms that would occur, it is natural to explore the possibility of applying the scheme of estimating unseen items based on existing data.

Putting technical details aside, the basic starting points of observation can be intuitively explained as follows (Kageura 2012). If it is supposed that:

1. We have a box containing possibly infinite number of fruits and we do not know how many types of fruits are in the box.
2. We take ten fruits out of the box.
3. We observe one of the following three patterns.
 - a. The ten fruits are all apple;
 - b. The ten fruits consist of three apples, two oranges, two bananas, one mango, one papaya and one kiwi fruit;
 - c. The ten fruits consist of ten different types of fruits.

What do we expect the eleventh fruit to be? In case 3-a, we would expect the eleventh fruit to be another apple and perhaps infer that the box only contains apples. In case 3-b, on the other hand, we would expect the eleventh fruit to be a new fruit type, though at the same time expect that it could as well be another apple, or other types already observed. In case 3-c, we would expect even more strongly than in case 3-b that the eleventh fruit would be a new type, and infer that the box contains a wide range of fruit types. These expectations and inferences can be justified by theories of probability and statistics (Good 1953).

If we can talk about the occurrence of new items based on the distributions of existing items in actual data, it is also possible to envisage terminological growth based on existing terminologies. Within the sphere of Terminology, given the terminology of a domain, it is the constituent elements of terms and not the terms themselves that have distributional information (Kageura 2002, 2012). Thus the basic framework for applying this to the phenomenon of terminological growth would be the following:

1. Assume that each constituent element of terms represents a single conceptual characteristic.
2. Estimate the growth and model the growth patterns of constituent elements based on the distribution of existing constituent elements.
3. Plug in the growth patterns to a relevant part of the terminological structure.

For instance, Kageura (2002) first observed how borrowed and native constituent elements are to be used in forming terms when a given terminology grows and expands. For each conceptual category represented by the terms, he then observed how new constituent elements were introduced when certain conceptual categories were deployed as heads or modifiers in the formation of a complex term. Figure 5 illustrates this latter framework. Figure 5 shows how it is possible to estimate which existing elements can be reused and how many new elements are to be introduced in constructing new terms in head and in modifier positions. This amounts to the extension of the classificatory patterns of term formation introduced in 4.1 at the most specific level of conceptual characteristics, i.e., those represented by individual constituent elements of terms.

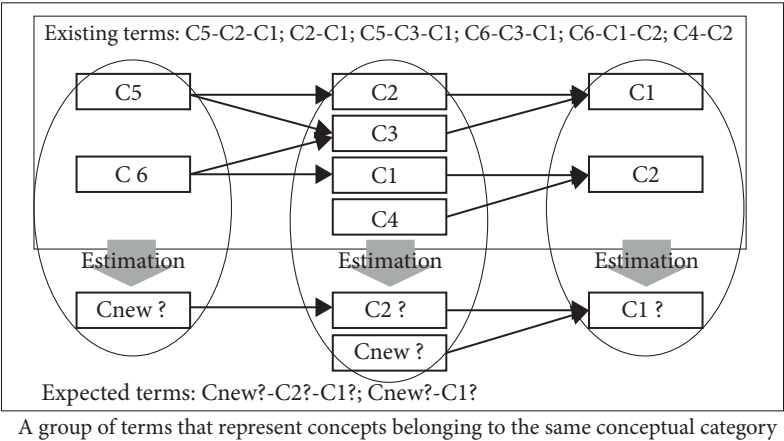


Figure 5. Terminological growth as captured by the growth of constituent elements

This framework can be further augmented in such a way that the introduction of new simple terms is taken into account. For this, the distribution of the length of new terms as reflected in the number of constituent elements should be incorporated into the model (Kageura 2002). As of now, to the best of our knowledge, no work has incorporated this factor in the study of terminological growth, except for the simplest assumption that the distribution of term lengths remains unchanged. Although some automatic term recognition methods incorporate the factor of length in evaluating termhood (e.g., Frantzi and Ananiadou 1999), it is only considered relative to a given set of candidates.

4.3 Growth patterns of modifications/conceptual specifications

As shown in Figure 5, the quantitative modelling of growth patterns of constituent elements introduced in 4.2 only deals with classificatory patterns of term formation, based on the constituent elements of terms, and leaves how they are combined, or the conceptual specification patterns, to the qualitative scheme. Although quantitative approaches are by no means better, it is natural to look for a way that covers both the growth of constituent elements and the growth of combinatorial patterns or modifications, especially given that the characteristics of constituent elements and of combinatorial patterns are two essential points of observation in complex term formation. This research line is still virtually untapped although some existing work shows how it could be further explored (Kageura 2002; Kageura and Chen, 2019).

A key observation that can lead to the exploration of conceptual specifications within the framework of terminological growth is to define the terminology graph or network. The terminology graph can be seen as an extension of the structure given by a set of terms in Figure 5. Given the terminology of a domain, it is defined as follows (Kageura 2012):

1. Take each constituent element of terms as a node or a vertex.
2. Add an edge between the vertices when they are used to make a term.

Kageura (2012) analysed the nature of the overall terminological structure based on the terminology graphs, but did not address the issue of growth. Furthermore, since the terminology graphs in Kageura (2012) were undirected, they fall short of taking into account conceptual specification patterns properly.

Kageura and Chen (2019) extended this framework by defining the directed graphs in which the edges are directed, and introduced two types of graphs, i.e. specification graphs and classification graphs. They correspond to conceptual specification patterns and classificatory patterns of terminology formation, respectively, introduced in Section 4.1. Figure 6 shows the specification graph and classification graph constructed from the following artificial example of a terminology (an artificial example is used because selecting examples sampled from real terminologies to show the graphs is difficult due to sparseness):

text classification, document classification, information retrieval, library classification, automatic text classification, document information, medical information, medical aid system, medical diagnosis, disease diagnosis, diagnosis record, disease image record

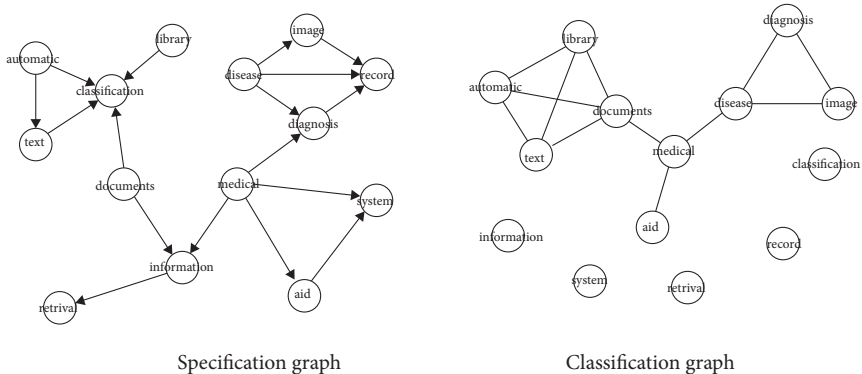


Figure 6. Specification graph and classification graph

As all the terms in a terminology are integrated into these graphs (although isolated terms that are not connected to other terms may well exist), these can be used as the basic starting point of modelling terminological growth. Although Kageura and Chen (2019) did not directly model the growth, they examined several dynamic network models by using entropic measures. Of particular importance from the viewpoint of modelling terminological growth is that the specification graphs may well be simulated by the preferential attachment model of network dynamics proposed by Barabási and Albert (1999). This gives the potential of describing the growth of specification graphs as a whole. The exploration to this direction is in its embryonic stage, and there are issues that still need to be solved. For instance, the representation of terminologies in specification and classification graphs do not explicitly indicate the level of terms. How the level of terms is to be incorporated into this modelling framework, should be examined in much greater depth.

4.4 Issues in the study of terminological growth

The study of terminological growth contains an inherent dilemma. On the one hand, it is essential to take into account the concreteness of terminologies, which cannot be addressed in the studies of term formation, at least not in sufficient depth. On the other hand, the study of growth inherently requires extrapolation to the actual or theoretical future state of terminologies, which by definition does not exist at the time when the study is carried out and thus necessarily remain a theoretical abstraction. In-between possible forms and those that actually exist, the concept of the possibility of existence had to be introduced as the target of research in terminological growth.

This signifies that there are a variety of practical issues that need to be solved. Firstly, as we indicated in Section 1, terminological data tend to be regarded as more artificial than textual data. This is less the case than many people seem to think. When we talk about the textual corpus of a domain, the choice of the range of texts always involves a substantial degree of arbitrariness. Even when people talk about a representative corpus, what it represents is already defined. Moreover, in terms of the knowledge and conceptual system of the domain, some of the research trends reflected in short-term textual productions can be ephemeral.

Nevertheless, it is more difficult to obtain suitable datasets that can be used for studying terminological growth, including the validation of the study both on a synchronic and diachronic scale. There are some well-maintained terminological databases such as SNOMED which can possibly be used for diachronic validations (cf. Bodenreider, Cornet and Vreeman 2018; Huser 2016), but their domains and languages are limited. Standard methods for evaluating the goodness of model in

the field of Natural Language Processing cannot be used for the task of evaluating the model of terminological growth. For instance, if we try to validate the structural model of terminological growth by using a method similar to the hold-out method,² which is the most widely used in the validation of NLP applications, the random choice of the hold-out data is inherently unsuitable for the validation of the model of terminological growth. This is because the study of terminological growth is concerned with concrete items (terms), and terminologies constitute a system. If we try to validate the model by using two terminological datasets for the same domain produced at different times (cf. Kageura 2002), the editorial policy and coverage of terminologies pose a problem for accurate evaluation. The terminological data that have been maintained under the same compilation policy and with time stamps for entries are limited. Using the set of terms extracted from textual corpora is not a solution, as textual corpora reflect research trends and also the arbitrariness of text choice.

In quantitative approaches, data size can also pose a problem. Sizes of terminology datasets are in general much smaller than the sizes of textual datasets for the corresponding subject field. For statistical estimation to give a valid or reliable result, data with a reasonable size are needed. If one tries to combine quantitative modelling with qualitative analyses, the size issue becomes even more critical. This is not a practical limitation that can be covered by more funding or efforts to enlarge the data because the size of the data is inherently limited by existing terminologies. This is indeed a theoretical limitation, as we cannot freely talk about the putative terminology of a given domain. It is thus important for research community not to automatically apply what is regarded as a norm in text-based methods to terminology-oriented studies.

Separately, the framework introduced in this section only covers complex term formation. Though this is reported to be the most frequent term formation method, there are other ways of forming terms, which cannot be ignored if we are to fully cover the phenomena of terminological growth, even if we limit our viewpoint to synchronic and structural growth. How to take into account other types of term formation within the framework examined here is virtually unaddressed.

2. The hold-out method is the method of evaluating a language model by using a part of language data reserved for evaluation. The other part is used for training the model (Manning and Schütze 1999; Eisenstein 2019).

5. Practical implications

The study of terminological growth starts from the recognition of the terminological sphere as the primary target of research (although this does not exclude taking into account other factors). This has several practical implications. Although not directly related to terminological growth itself, text-based processing, in its purest form, cannot distinguish bias from proper statements or concepts, as was vividly and painfully revealed by Tay, the Microsoft chatbot that made its debut on Twitter in March 2016 and became a fascist-misogynist within 24 hours (Hunt 2016). One of the ways to counter this is to refer to a system of concepts within which consistencies can be validated. This has been socially supported by dictionaries and terminologies (cf. Kageura 2020).

Although in specialised domains we can safely assume that this issue of bias is less serious, identifying terminologies as an independent sphere is important not only theoretically but also practically. For instance, in practical translation, language service providers (LSPs) often maintain their own in-house domain terminologies that are not attributed to a particular group of documents that they translate (cf. Kageura and Marshman 2019). For automatic term extraction to contribute to the management of these terminologies, the extracted terms should be evaluated in relation to existing terms in the terminologies rather than an *in vitro* gold standard. If the range of terms that can be possibly created has been estimated in the study of terminological growth, there is a possibility that the target of automatic term recognition can be narrowed down to fit the expectation of existing terminologies and the task of maintaining these terminologies.

Chen and Kageura (2020) show that the incorporation of terminological structure into neural bilingual term induction methods improves the performance of bilingual induction. As information on terminological structure is, in a way, used as an estimation of proper terms in the target language, this can be regarded as a contribution to information on terminological growth in a practical application. In automatic term extraction, filtering of results has become an important research area (Drouin et al. 2018). The estimated range of terminological growth can be used for filtering purposes. These schemes can be applied to terminology planning as well (cf. Zarnikhi 2014). Although the level of application is different, estimation of the range of terms can be used for evaluating the sufficiency of textual evidence in collecting terms (Miyata and Kageura 2018).

As research in terminological growth itself is in its very early stage, actual cases of practical applications are still limited. However, these cases show that there are areas in which such research can make a practical contribution.

6. Conclusions

Terms and terminologies are concrete substances, and the essential *raison d'être* of terminology as an independent research field or programme resides in addressing this concreteness. As such, the core part of terminology is closer to lexicography than to lexicology. While achievements in studies of term formation are laudable in themselves and in relation to such applications as automatic term recognition, as a theoretical programme in terminology, the study of terminological growth has remained an important research topic that has not been addressed to any significant extent.

This chapter began by consolidating the sphere of Terminology and the essential research question that the study of terminological growth should have at its core. Based on this research question, the current state of synchronic/structural approaches to terminological growth was summarised, together with some theoretical and methodological issues. Practical implications were also examined.

The problem definition given in the form of research Question (II) only covers a part of what can fall under the study in terminological growth in its wider sense. Though limited, what was elaborated on here is perhaps the first endeavour to address the issue of terminological growth in the sphere of Terminology, and thus of theoretical importance not only in terms of the technical approaches but also in terms of the problem setup itself. Within this limited framework, however, a number of issues remain unsolved, as observed in 4.4. Even after fully developing the modelling methodologies of terminological growth in structural and synchronic framework, the task of integrating it with or clarifying the relation to other related studies including diachronic approaches remains. This is a topic still in its embryonic stage and thus is a rich area to be explored.

Funding

This work is partly supported by JSPS KAKENHI Grant Number 19H05660.

Terminology and equivalence

Pilar León-Araúz

University of Granada

Equivalence has not attracted as much interest as other issues in Terminology. Nevertheless, terminological equivalence can be understood from different perspectives: (1) equivalence at the concept, term or textual level; (2) degree of equivalence (from full to zero equivalence); (3) interlinguistic or intralinguistic equivalence; (4) equivalence based on the syntactic structure of terminological correspondences; (5) equivalence depending on the solution found for each equivalence-related problem; or (6) discovered or created equivalence. This chapter addresses equivalence from different theoretical approaches and their impact on the practice of terminography by analyzing different equivalence-related problems, terminographic methods and products.

Keywords: equivalence types, degrees of equivalence, translation, terminology, variation, terminographic resources

1. Introduction

Equivalence has always attracted more interest in Translation Studies than in Terminology, probably because classical monoreferential postulates did not acknowledge problems of equivalence (Sager 1994b; Van Campenhoudt 2001). This is only natural since translation is the process in which a source text is rendered into an equivalent target text, first in terms of meaning and secondly in terms of style (Nida and Taber 1969). However, as opposed to text-based approaches, terminographers usually conceive equivalence at the concept/term level, because their goal is to situate concepts and terms in terminographic resources.

Equivalence became an issue in Terminology when its theory and practice began to develop in descriptive and applied scenarios (i.e., translation). That was when specialized communication stopped being the exclusive domain of subject field experts, and variation (Freixa; Dury; Fernández-Silva, chapters in this volume) came to the foreground. Equivalence and variation became inseparable from the moment that different languages, cultures, and communicative situations were collectively

taken into consideration. It is term, cultural, and conceptual variation that triggers the different degrees of equivalence between a source and a target concept/term, or even between two concepts/terms within the same language.

For instance, in cross-lingual scenarios, conceptual anisomorphism may lead to a lack of conceptual equivalence (e.g., *crime* and *delito* in Spanish, since what constitutes a *crime* or a *delito* can vary, depending on the jurisdiction), whereas pragmatic variation may cause partial term equivalence, even if equivalence at the concept level is fully achieved. This is the case of the formal Spanish variant, *can* as opposed to the less formal term, *perro*, both of which mean *dog*. Since translation is one of the ultimate purposes of multilingual terminographic resources, the notion of terminological equivalence cannot disregard translation equivalence, although their scope and practical challenges are not exactly the same.

This chapter addresses equivalence from different theoretical approaches and their impact on the practice of terminography through the analysis of different equivalence-related problems, terminographic methods, and products. The remainder of this chapter is organized as follows: Section 2 discusses the notion of equivalence in Translation and Terminology; Section 3 explores the different types of equivalence considered in Terminology; Section 4 describes how equivalence is approached in terminographic resources; Section 5 summarizes the content of the chapter and presents conclusions.

2. Equivalence in Translation and Terminology: Differences and common ground

Equivalence has long held a privileged though controversial position in Translation Studies since authors define it differently, based on formal, semantic, pragmatic, and/or textual criteria. Apart from the traditional opposition of faithful/free, other pairs of terms, such as semantic/communicative (Newmark 1981), formal/dynamic (later called functional) (Nida and Taber 1969; Nord 1997), direct/oblique (Vinay and Darbelnet 1995), and textual/pragmatic (Baker 1992) have been proposed as descriptions for the degree of perceived similarity at the level of form and/or function between a source text and a target text.

However, these changes of label have not always been accompanied by significant new insights into the nature of equivalence. In fact, Baker (1992, 5–6) regards equivalence as a convenient and familiar notion with no theoretical status; Snell-Hornby (1988, 22) defines it as an imprecise ill-defined term presenting the illusion of symmetry between languages; and Pym (2011, 81) considers it to be a relation that many believe in and which no one can prove beyond the level of

terminology, since it has no ontological foundation. Part of this complexity is due to the fact that the rules do not remain the same, but change with each new translation context (León-Araúz and Faber 2014). This complicates things for systematization purposes, which is the aim of terminographic resources.

While translators look for equivalence at the sentence or text level within a particular translation commission, terminographers seek decontextualized equivalence between concept/term entries (i.e., when designing terminographic resources). The notion of equivalence in Terminology is thus closer to that of bilingual lexicography (regardless of the diverging onomasiological/semasiological approaches). In fact, Atkins (2008, 44) acknowledges that there is considerable difference between the context-sensitive equivalence sought by the translator and the context-free equivalence that must be offered in a bilingual dictionary entry.

Along these lines, Kerremans and Temmerman (2016) state that translators are not always guided by the principle (often adopted for structuring multilingual term bases) that a term in the source language should be rendered as a direct (literal) equivalent in the target text. Likewise, Rogers (2008, 102) points out that the focus in Translation studies is clearly on text whereas in Terminology, it has been on concept systems.

Therefore, terminological equivalence does not always correspond to translation equivalence (Kerremans and Temmerman 2016, 59). It is thus a broader concept for translators than for terminographers. Consequently, equivalents at the sentence or text level may be those reproducing the same function or effect than the source text rather than just those conveying the same concept. The notion of translation-oriented equivalence allows for a wide spectrum of translation mechanisms (e.g., hypernyms in the target language, paraphrastic structures, circumlocutions, omissions, modulation, transposition, etc.) that are rarely if ever stored in term bases. This is why translators often resort to “unstructured resources”, such as texts originally written in the source and target languages (comparable corpora) or previously translated texts (parallel corpora) (Kerremans 2017).

For these reasons, Arntz (1993) acknowledges that the conditions for interlingual comparison are more favorable in the field of Terminology, where the conceptual content of terms is most important. Yet a uniform conceptual basis straddling language boundaries cannot be assumed (Arntz 1993, 6), and establishing equivalence between two specialized units is not always as straightforward a task as it may seem at first (L’Homme 2020). In fact, common cross-lingual equivalence-related problems in both Translation and Terminology arise in regard of concept and term dynamics (León-Araúz 2017).

3. Equivalence types in Terminology

The main debates on equivalence in Terminology revolve around the scope of equivalence, degrees of equivalence, and equivalent formation devices, which give rise to different classifications.

3.1 The scope of equivalence

In Terminology, equivalence was initially defined as the relation between concepts sharing the same intensional identity (i.e., having the same features and position in the concept system of a particular subject field) (Arntz and Picht 1989b). Equivalence was thus established exclusively at the concept level. However, the ISO standard 1087-1:2000 regards equivalence at the term level: “the relation between designations in different languages representing the same concept”.

Likewise, in the *Glossary of Terms used in Terminology* (De Bessé, Nkwenti-Azeh and Sager 1997, 131–132), equivalence is defined as “the relationship of formal, semantic, or pragmatic correspondence between two linguistic units”, whereas *equivalent* has two senses: (i) “an expression in one language which has the same sense and reference as an expression in another language” and (ii) “synonym: a word or term which has the same denotation or the same sense as another”. For Dubuc (2002, 55), two terms are equivalent if they share full sense and usage identity within the same domain. These last definitions seem to match that of equivalent in bilingual lexicography as observed by Zgusta (1971, 312): “a lexical unit of the target language which has the same lexical meaning as the respective lexical unit of the source language”.

The opposition concept/term-based equivalence goes hand in hand with the opposing views of knowledge-driven and lexicon-driven approaches (L’Homme 2020), but the previous definitions suggest three intertwined interpretations (Figure 1). In the first case, equivalence is sought among concepts, with no reference to the terms, as they are considered language-independent. Concepts A and B would actually be the same concept. Since terms act as pointers to concepts, the second interpretation still accounts for the concept as the central anchor of equivalence, even though the relation is established between the designations.

In this sense, Rondeau’s definition (1984) would be a mixture of both cases: two terms are considered equivalents if the designation in one language shares an identical relation with the designation in the other language and if the concept in one language shares an identical relation with the concept in the other language. In the third case, however, other factors are taken into account, such as pragmatics, sense, reference, usage, and lexical meaning. Therefore, equivalence can also be

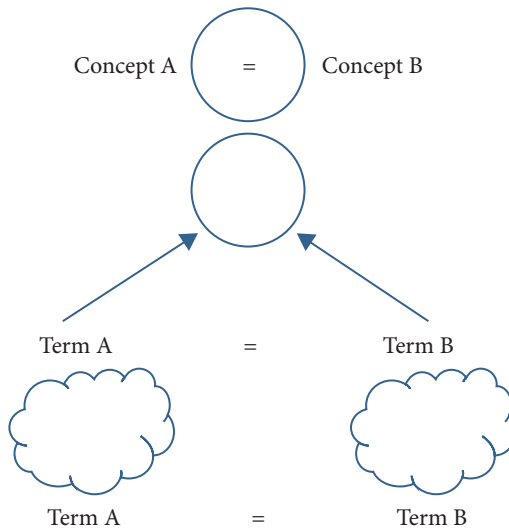


Figure 1. Concept/term-based equivalence

established between terms on the basis of their meaning rather than according to their potential to label a predefined concept (L’Homme 2020, 231). These distinctions, subtle as they might seem, may impact the way terminographic resources are built (Section 4).

Nevertheless, equivalence may go beyond concepts and terms. As stated by Rogers (2007a, 2), this view of equivalence still only focuses on denotation: “what happens when a term is plucked from a dictionary and inserted into a text?” Text-based equivalence would thus be closer to the notion of equivalence in Translation Studies, but it would also include other specialized units of interest in Terminology, such as phraseology and verbal structures.

3.2 Degrees of equivalence

In any case, whether considered at the concept, term, or textual level, the complexity of equivalence is generally acknowledged to be a matter of degree. A common classification of equivalence based on this criterion is shown in Figure 2 (Arntz 1993; Felber 1987), where the degree of equivalence depends on the degree of congruence between the systems of concepts in different languages (Picht and Draskau 1985). For the sake of simplicity, equivalence-related problems are usually grouped into four degrees of equivalence.

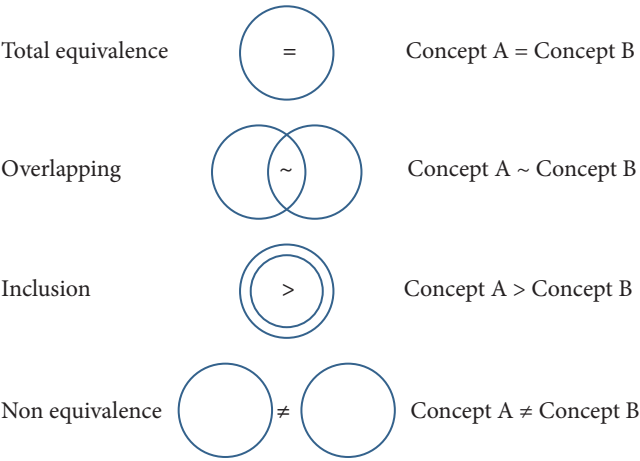


Figure 2. Types of equivalence based on degree (adapted from Felber 1987)

1. Total conceptual equivalence: The content of terms in different languages point to concepts that occupy the same position in all systems (i.e., they are related with the same number and types of concepts in the same hierarchical levels) (e.g., *leukocyte* and the Spanish *leucocito*). Regarding the number of terms representing the concept in two languages, the following cases are possible: one:one, one:many and many:many (Arntz 1993), thus accounting for synonymy.
2. Conceptual overlapping: Two terms point to a common conceptual content, but also add other diverging features (e.g., *pier* and the Spanish *embarcadero*, which operate on different design principles and have a different cultural role).
3. Inclusion: One term points to a concept that is included in the concept pointed by the other term, the latter including further features, which means that one concept merges into another (e.g., the French *fleuve* and the Spanish *río*, which encompasses both *fleuve* and *rivière* because no distinctions are made).
4. Non-equivalence: There is not a common concept lexicalized in source and target languages, usually due to culture-specific terms, neonyms or other lexical gaps. It may occur that the concept only exists in a language culture (e.g., *pejerrey*), but also that the concept exists in both cultures but only one has a designation for it. Legal terminology has been very often used to illustrate non-equivalence (Sandrini 1996; Šarčević 1997; Pimentel 2012).

There are other classifications based on degrees of equivalence, ranging from two to up to five degrees, from which the simplest would be that of three degrees: full, partial, and non-equivalence. For instance, the ISO standard 25964–1:2011 also

considers four degrees of equivalence, yet between terms (exact, inexact, partial and non-equivalence). Yong and Peng (2007) distinguish five types within partial equivalence: approximate, analytical, synthetic, subordinate, and superordinate equivalence. In contrast, Van Campenhoudt (2001) distinguishes two more types within inclusion (i.e., inclusion and hyponymy and inclusion and meronymy), whereas Arntz (1993, 14) distinguishes two more types within overlapping, i.e., cases where the intersection is large enough to establish equivalence and cases where it is too small.

The problem is, according to Sandrini (1996), that there is no criterion for the terminologist to decide whether there are enough overlapping characteristics to justify a common concept. In some cases, one differing feature could lead to two different concepts. In other cases, quite a few differing features are not sufficient to abandon equivalence. As a result, Sandrini (1996) prefers to consider only two main types of equivalence: absolute or total equivalence and partial, relative or near equivalence. Šarčević (1997, 237–239), in contrast, argues against the existence of absolute equivalence and proposes three types: near equivalence, partial equivalence, and non-equivalence. Finally, the International Terminology Working Group (1997) proposes up to five types: exact equivalence, inexact equivalence, partial equivalence, non-equivalence, and a new type that they call single-to-multiple equivalence, for concepts in a source language that are not recognized as a single term in the target language.

These classifications are usually considered in interlingual comparisons, but all of these degrees, except for non-equivalence, can also occur in intralingual scenarios, since synonyms, or term variants, can also codify varying degrees of concept or term equivalence. Although the ISO standard 25964–1:2011 considers that synonymy is not within the scope of equivalence, we agree with De Bessé, Nkwenti-Azeh and Sager (1997) who include synonym as a second sense of equivalent. After all, equivalence can be understood as a broad spectrum of relations, from similarity to identity (Adamska-Sałaciak 2010, 403) and the ISO standard 1087 (2000) defines synonymy as the relation between or among terms in a given language representing the same concept.

Therefore, synonymy and equivalence in ISO standards are the same phenomenon except for the number of languages involved, even though the degree of similarity sufficient to regard two words as intralingual synonyms is typically lower than that necessary to regard them as interlingual lexical equivalents (Adamska-Sałaciak 2013). Furthermore, when terminographic resources include large collections of term variants in their entries, degrees of equivalence should not only be established with regard to the concept designated, but also among the terms themselves, both interlinguistically and intralinguistically (Figure 3).

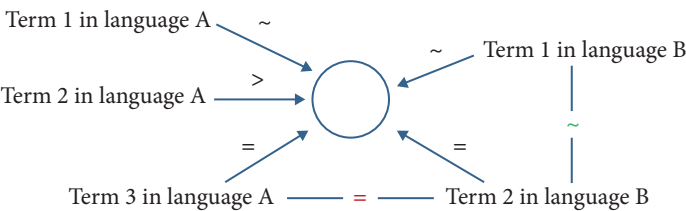


Figure 3. Term variation degrees

For instance, the terms designating CLIMATE CHANGE in English (*climate change*), Spanish (*cambio climático*) and French (*changement climatique*), show total equivalence at the concept level. In a descriptive resource, other designations could also be included at the term level, ranging from slight morphosyntactic changes with regards to the main term (*changement du climat*, *changement de climat*, as reflected in TermiumPlus) to larger deviations from a conceptual point of view (*dérèglement climatique*, *emergencia climática* or even *global warming*). Strictly speaking, however, *global warming* does not refer to the same concept but is often used interchangeably because one is part of the other. It would thus fall under the inclusion type of equivalence. *Dérèglement climatique* and *emergencia climática* show a change in the head of the compound reflecting multidimensionality (Bowker, this volume) and adding new semantic traits to the same phenomenon. Consequently, interlingually, *global warming* would be totally equivalent to *calentamiento global* but both would be partially equivalent to the canonical concept designated by *climate change*. Intralingually, *changement climatique* and *changement du climat* would be totally equivalent but they would hold a partial equivalence relation with *dérèglement climatique*.

The degree of equivalence can thus be linked to the degree of variation, which in turn depends on the causes underlying term variation. According to Freixa (2006, this volume), causes can be (i) dialectal (based on geographical, chronological or social origin); (ii) functional (field, tenor and channel); (iii) discursive (related to text cohesion); (iv) interlinguistic (when two or more languages are in contact, through borrowings and calques); and (v) cognitive (based on the different conceptualizations of reality). Although cognitive variants, such as *emergencia climática* or *calentamiento global*, are the main source of partial equivalence at the concept level, other factors (e.g., dialect, register, level of specialization) also challenge equivalence at the term level and thus determine interchangeability in a text.

In this sense, equivalence among terms can also be studied from a structural point of view (Gaussier 2001; Tran, Burgun and Garcelon 2003), especially with regards to multiword terms and term variation. For instance, *air pollution* (N+N) can be rendered into Spanish as *contaminación del aire* (N+Prep+N) or *contaminación atmosférica* (N+Adj). At the concept level, both terms would be total equivalents,

but morphosyntactically the first would be ‘more equivalent’ to the source term. Nevertheless, structural equivalence is not necessarily a good criterion for the choice of the best equivalent, since other factors, such as frequency or pragmatics are often more relevant. Knowledge-driven perspectives have given little importance to the structure of terms when establishing equivalence (L’Homme 2020), even though the linguistic structure of concepts sometimes impedes the recognition of the conceptual structure of a subject field, which is the basis for the satisfactory matching of equivalents (Sager 1994b). For example, two terms sharing the same head and the same number of constituents do not necessarily point to two sibling concepts. This is the case of *water turbine* and *cross-flow turbine*, where the second is actually a type of the first. Although the concept also adopts the form *cross-flow water turbine* (pointing to the “right” hierarchical level), the omission of formants is very frequent in multi-word terms (MWTs) and can be misleading while establishing equivalence.

3.3 Equivalent formation devices

Equivalence can also be categorized based on the solutions given to varying degrees of equivalence. Widely accepted classifications in bilingual lexicography include the following: (i) cognitive (semantic, systemic, prototypical conceptual, decontextualized, notional) equivalence; (ii) translational (insertable, textual, contextual) equivalence; (iii) explanatory (descriptive) equivalence; (iv) functional (situational, communicative, discourse, dynamic) equivalence (Zgusta 1971; Adamska-Sałaciak 2011). In the context of terminographic resource localization León-Araúz and Faber (2014) proposed the following classification.

1. Canonical equivalence applies when no equivalence problems arise (i.e. total equivalence) and the translation relation may be symmetric. *River* and *río* are thus canonical symmetric equivalents, but this does not mean that when canonical equivalents are found no other equivalents are possible, since context can modify the degree of equivalence.
2. Generic-specific equivalence addresses problems related to cross-lingual categorization differences (i.e. inclusion), depending on the communicative situation and directionality. A specific-generic equivalence applies when translating the term *shingle* (an umbrella concept unnamed in Spanish covering both *sand* and *gravel*), which in Spanish can be translated by its hypernym *material de grano grueso* (*coarse material*). In the same way, when the context describes a beach nourishment scenario in Spanish, *material de grano grueso* can be translated not only by its canonical form *coarse material*, but also by its specific equivalence *shingle*. Alternatively, the following equivalence may apply.

3. Extensional equivalence is a kind of generic-specific equivalence, because the original term is translated by all of the hyponyms of the concept in the target culture. In this way, *shingle* can also be translated by the enumeration of its subtypes (*arena* and *grava*).
4. Metonymic equivalence applies when original terms are expressed in the form of a metonymic variant and target terms are not. *Groyne* could be translated into Spanish both as *espigón* or *escollera* (rubble mound, metonymic variant). However, *escollera*, in its coastal structure sense, can only be translated as *groyne*.
5. Communicative equivalence applies when the concept exists in both language cultures but only in one language the concept has been lexicalized in several variants with different communicative or conceptual consequences. It thus addresses equivalence at the term level and involves establishing register correspondence between domain-specific and diaphasic variants. The canonical equivalent of *lodo* is usually *mud*, but in the domain of water treatment, experts prefer *sludge*. Furthermore, depending on the communicative situation, certain terms can be translated as the expert neutral variant or the lay-user variant in the target language (e.g., *intestines* or *bowels* for *intestinos*).
6. Functional equivalence applies when cross-cultural differences are an obstacle to translation and affect both concepts and terms (i.e. overlapping). It involves deculturalising original terms so that receivers can relate to the concept. *Muskeg* can be translated as *turbera*, and *malecón* as *seawall*. Even though the translation equivalents lose the cultural meaning of the original term, they designate the closest concepts in target cultures from a semantic perspective. Other terms, such as *quay*, *dock*, and *wharf* must rely on additional contextual features, since they can all be translated into Spanish as *muelle*, *embarcadero* and/or *dársena* depending on the size, function, and position of the structures. This equivalence relation is particularly asymmetric.
7. Cultural equivalence is another way of addressing cross-cultural differences and consists of adapting original culture-bound terms to other culture-bound terms in the target culture. For example, the usual canonical translation of *pier* is *embarcadero*, but piers are often recreational areas that do not correspond to the Spanish concept. In these cases, the most suitable translation is *paseo marítimo* [boardwalk], or even *malecón* or *costanera* [seawall] for South American Spanish since, even though these kinds of construction are slightly different, the cultural component of the concept is preserved.
8. Descriptive equivalence also addresses culture-bound problems and makes explicit certain semantic features depending on user communication needs, often to distinguish a concept that has not been designated in the target culture. For lay users, for example, *muskeg* could be translated as *humedal canadiense*

muskeg [Canadian wetland muskeg], which involves adding its hypernym and location. In contrast, the term *jetty* can be translated into Spanish as *espigón*, which is the canonical translation of *groyne*. Another possibility is *dique*, a functional translation, based on the wide array of functions it may have. However, if both terms are found in a text (*jetty* and *groyne*), some distinction should be made. In this sense, a descriptive translation could be *espigón de encauzamiento* (literally *channeling groyne*), which explains the function of jetties and distinguishes it from groynes.

9. Non-translation equivalence, or loans, not only addresses culture-bound problems when concepts and/or lexicalizations do not exist in the target culture (e.g., *pejerrey*, a fish that only exists in South America), but also cases in which descriptive equivalence is not appropriate. For example, terms like *muskeg* or *billabong* can be kept in their original form if the receivers are experts who do not need any description.
10. Neological equivalence addresses non-equivalence when descriptive or non-translation equivalence is discarded, by employing calques or neonyms.

Generally speaking, total equivalence is addressed with canonical equivalence and, depending on context, inclusion corresponds to generic-specific, extensional or metonymic equivalence. Overlapping can be addressed with communicative, functional or cultural equivalence and non-equivalence with descriptive, non-translation or neological equivalence.

All of these solutions can in turn be classified into two broader categories based on whether equivalence is discovered or created (Adamska-Sałaciak 2010). Sager distinguishes genuine terms from translation equivalents. He defines translation equivalents as substitute lexical items or expressions in a target language for a source-language lexical unit that has no matching concept in the target language culture (Sager 1994b). Similarly, Adamska-Sałaciak (2010) and Pym (2007) make the same distinction with the opposition natural/directional equivalence.

Consequently, even though authors such as Arntz (1993) only envisage terminological equivalence as conceptual equivalence, terminological equivalence has different types: (i) equivalence at the concept, term or textual level; (ii) degree of equivalence (from full to zero equivalence); (iii) interlinguistic or intralinguistic equivalence; (iv) equivalence based on the syntactic structure of terminological correspondences; (v) equivalence depending on the solution found for each equivalence-related problem; or (vi) discovered or created equivalence. All of these factors can give rise to complex typologies (e.g., partial conceptual but full pragmatic equivalence or vice versa), which determine how equivalence is understood by terminographers and reflected in terminographic resources.

4. Equivalence in terminographic resources

The assumptions discussed in previous sections have a direct impact on the design and population of terminographic resources, especially in multilingual settings because of how different languages subdivide reality (Arntz 1993).

The representation of multilingual equivalence highlights the consequences of conceptual anisomorphism and polysemy. It also requires a reflection on meaning, sense and concept, similarity or interchangeability. In Terminology, such notions are often fuzzy depending on whether the focus is on the concept or the term, and whether there is one pivot language or many languages simultaneously.

Cabré and Tebé (2005) distinguish multilingual terminological work from monolingual terminological work with equivalences. In multilingual work, all languages are equally represented. Terms in all languages are independently researched and only after that, are correspondences in different languages established. In contrast, monolingual work focuses on one language, and equivalent designations are added, with no additional information.

Theoretically, the first choice is favored in Terminology, since it accounts for a concept-oriented approach and a more representative view of linguistic diversity. Nevertheless, in practice this is not always the case, even for resources that were initially created with this approach in mind. In fact, the POINTER (2011) report argues that the more languages multilingual resources include, the lower the quality of the resource, especially with regards to languages other than the source language. In many cases, there is a tendency merely to translate terms when adding equivalents which, together with the lack of conceptual organization, results in a high percentage of irrelevant terms (POINTER 2011). POINTER recommendations include ensuring that equivalents are based on a terminological rather than a lexicographical point of view.

A distinction is thus made between correspondence and equivalence (i.e., natural or directional equivalence, Section 3), which can be linked to the opposition of comparable versus parallel corpora when researching terms in the target language. However, the truth is that both types of term coexist in most terminographic resources, especially if the resource is aimed at translators. Sager (1990, 224) recognizes that a number of terms may only exist as translation equivalents (e.g., the names of institutions of a particular country), but even in cases where genuine correspondences are found, terminographers might also want to store other translation equivalents within the same record. In fact, Kerremans and Temmerman (2016) argue in favor of including 'deviations' from standard equivalents (i.e., found in parallel corpora) in terminographic resources. As a result, in the same resource, there might be entries both with correspondences and equivalents, others with only correspondences, and others with only equivalents. The challenge lies in documenting

the particular uses of each option. In the context of bilingual lexicography, Atkins and Rundell (2008, 45) state that when space cannot be spared for detailed meta-linguistic indications, the lexicographers must choose a safe general translation in preference to one that may be perfect in some contexts but perhaps erroneous in others. However, in term bases, space is not a problem.

Consequently, the approach to equivalence influences the different methods employed in terminographic practice and its final products, especially regarding the way in which data are interpreted to establish equivalence (Section 4.1) and how it may be represented in the resources (Sections 4.2).

4.1 Establishing equivalence

According to Picht and Draskau (1985), the assessment of equivalence is one of the main tasks of terminological analysis. However, the way in which equivalence should be established is not entirely clear (Pimentel 2012). This is probably because not all terminographic products are the result of an entirely onomasiological or semasiological approach, nor do they all have the same purpose (i.e., translation, standardization), and not all equivalence-related problems can be solved in the same manner (Section 3.3).

Hartmann (2007, 15) argues that it is usually a patchy, partial, intuitive, arbitrary and tentative process of limited generalizability. In contrast, Adamska-Sałaciak (2010, 403) considers that equivalence judgements – like all similarity judgements – are creative processes in the mind of the equivalent-seeker. Nevertheless, several methodological approaches to the identification and validation of equivalences have been proposed both in Terminology and bilingual lexicography.

From a knowledge-driven perspective, the assessment of equivalence is based on congruence with a concept system. According to Arntz (1993), a concept can only be understood in the context of the system to which it belongs. Concept systems have been defined as semantic networks that allow the aggregation of multilingual equivalents of terms and monolingual terms that are synonyms into a common concept object (Shreve 1995). Thus, a comparison of two languages first requires the mapping of the concept system pertaining to each individual language (Arntz 1993). Nuopponen (2018b) admits that an equally extensive analysis in all languages can be very time-consuming, and that it is usual to focus on the main language/languages of the project and apply systematic methods on it/them whereas other equivalents are compiled in a more or less ad hoc manner.

From a lexicon-driven perspective, Adamska-Sałaciak (2010) states that equivalence can be established based on one of the following features: import, meaning, importance, significance, value, force, power or efficacy. On a more practical level, Atkins and Rundell (2008) suggest using comparable and parallel corpora to

establish equivalence based on the following elements: (i) semantic content (denotation and connotation); (ii) collocational context (equivalence may vary based on collocations); (iii) vocabulary type (register, style); (iv) message; and (v) function.

Sager (1994b) acknowledges that only with detailed analyses of both knowledge structures in two cultures and their linguistic representations is it possible to determine the degrees of coincidence between two terminological systems. He thus recommends a dual analysis in which concepts and terms are equally important. Reconciling knowledge-driven and lexicon-driven perspectives, Pimentel (2015) argues that the assignment of equivalents should be based on (i) corpora; (ii) analysis of the semantic content/conceptual characteristics; (iii) analysis of the collocational context; (iv) encyclopedic knowledge; and (v) a *tertium comparationis* external to terms based on a situational and a formal dimension, which Pimentel (2012) subsumes under the notion of frame.

Ideally, the independent research of terms in each language involves the identification of concepts and relations with a view to analyzing and representing them in concept systems. Only after concepts are defined can conceptual correspondences be established. This means that terms are not 'translated', and that correspondences emerge from independent concept analysis. This inevitably leads to different degrees of equivalence and different sets of term variants. It is then necessary to assess the degrees of equivalence and decide how overlapping, inclusion, and non-equivalence should be dealt with. Finally, usage-related information must also be collected in order to relate variants and discriminate their usage.

Conceptual structures can be discovered or verified through corpus-based or definitional analysis. In the first case, concepts can be searched for, in comparable corpora, together with knowledge patterns (KPs) (Meyer 2001; Marshman, this volume) in order to find the explicit occurrence of semantic relations and thus reconstruct the conceptual system. In the second case, definitions in different languages are compared to ascertain that the main conceptual features match. For instance, in Figures 4 and 5 the terms *boundary layer* and its Spanish translation, *capa límite*, are surrounded by hyponymic KPs (e.g., *which is, known as, called, can be defined as, is, llama, (, -, es) that*) that relate both terms to similar hypernyms (e.g., LAYER, REGION, PART) and other concepts (e.g., FLUID, ATMOSPHERE). The same hyponyms of BOUNDARY LAYER are found in both corpora (PLANETARY BOUNDARY LAYER and ATMOSPHERIC BOUNDARY LAYER). In Spanish, *capa limitrofe* appears as a term variant.

In Figures 6 and 7, the terms are surrounded by meronymic patterns (*is the part, composed, comprising, parte, constituye*) linking the hyponym PLANETARY BOUNDARY LAYER to the TROPOSPHERE in both languages. In the definitions in Table 1 both concepts show similar characteristics.

Earth's atmosphere. But there is something called the "Planetary **boundary layer** " which is the lowest part of the Troposphere (where we live and manages to break through what is known as the ' planetary **boundary layer** '. This layer is technically defined as the region of the atmosphere directly above Earth's surface, known as the atmospheric **boundary layer** (ABL). The effects of the surface – heating, cooling, and friction from backscatter lidar and radiosonde instruments. The Planetary **boundary layer** (PBL) is the lowest part of the troposphere which is directly above the Earth's surface, as well as golfing opportunities. The Planetary **boundary layer** (PBL) can be defined as that part of the atmosphere that is trapped within a very thin layer of the atmosphere called the **boundary layer** (BL). A hydrostatic, biaxial, vertically-pointing Mie Scatterer immediately above the earth's surface known as the surface **boundary layer** (or simply the surface layer). This layer is only a part of the troposphere, as air flows outward and mixes with air rushing in from the inflow **boundary layer** (that is, the layer of air near ground level). The combined flow of the atmosphere's lowest layer, whose thickness varies from a few meters to several kilometers, is called the **boundary layer** . Researchers study the data to learn more about the structure and dynamics of the lowest 1-2km of the atmosphere (**boundary layer**). Traditional climate research involved microclimate studies in the **boundary layer** , defined as the layer of air within a few meters of the surface of the airfoil section and on 2) The occurrences occurring at the **boundary layer** , it is basically a thin layer of air attached to a solid surface. A prominent example of such an application is the atmospheric **boundary layer** , the lowest layer of the atmosphere, in which many physical processes occur, such as low level winds and turbulence in the planetary **boundary layer** , which is the portion of the atmosphere closest to and most affected by the surface. I'm not into that style of music. Sammanfattning : The **boundary layer** is the part of the atmosphere that is in direct contact with the surface. Methods for some of the candidate alternative grids. See here . The **boundary layer** is the layer of the atmosphere in which we live and ranges from a few meters to several kilometers. In physics and fluid mechanics, a **boundary layer** is that layer of fluid in the immediate vicinity of a bounding surface where the velocity of the fluid is zero or close to zero. Understorms, frosts, fog and severe wind storms. The atmospheric **boundary layer** is the region adjacent to the surface in which there are significant temperature and moisture gradients.

Figure 4. *Boundary layer* and hyponymic KPs

el cuerpo está rodeado de una capa de aire, a la que se le llama **capa límite** , que puede adelgazarse rápidamente a causa del viento, pero también abarca los cambios climáticos del pasado. **Capa límite** : En términos generales, una capa de aire justo encima de la superficie terrestre. La **capa límite** atmosférica (la capa inferior de la atmósfera en interacción directa con la superficie terrestre), también conocida como la **capa límite** atmosférica (CLA), es la parte más baja de la atmósfera y su espesor varía entre los 10 y 100 metros. La **capa límite** es el área más cercana a la superficie donde las condiciones climáticas son más variables. La **capa límite** es la capa de aire más baja en la atmósfera terrestre (hasta 100 metros). Según la Tercera Ley de Newton. **Capa límite** (Boundary layer): La **capa límite** es una zona existente alrededor de un cuerpo que se desliza sobre una superficie viscosa. En la atmósfera terrestre, la **capa límite** es la capa de aire cercana al suelo y que se ve afectada por la fricción. En las condiciones iniciales (u0). En mecánica de fluidos, la **capa límite** o capa fronteriza de un fluido es la zona donde el movimiento del fluido cambia de manera variable, entre los valores de 10 a 100 metros. **Capa límite** planetaria: Capa atmosférica que se extiende desde la superficie terrestre hasta la tropopausa. **Capa límite** planetaria - la parte más baja de la atmósfera que llega hasta la tropopausa.

Figure 5. *Capa límite* and hyponymic KPs

Left context	KWIC	Right context
...ogen , the lightest element. The planetary boundary layer is the part of the troposphere that is nearest to the Earth's surface.	boundary layer is the part	of the troposphere that is nearest to the Earth's surface.
...oil-water interphase. The upper part of the boundary layer , composed of the previously described fluid boundary layer , is constant throughout. The air above the planetary boundary layer , comprising about 95 percent of the total mass of the atmosphere , is in the free atmosphere.	boundary layer , composed	of the previously described fluid boundary layer , is constant throughout.
...surface boundary layer. The ocean surface boundary layer is the part of the ocean that is directly exposed to the atmosphere.	boundary layer is the part	of the ocean that is directly exposed to the atmosphere.

Figure 6. *Boundary layer* and meronymic KPs

mejorar su enunciado. **Capa límite** atmosférica es aquella **parte** de la tropósfera directamente adyacente a la superficie terrestre. La **capa límite** planetaria, la **parte** más baja de la atmósfera de la Tierra, se conoce como capa superficial, representando la zona de mayor actividad meteorológica. La **parte** inferior de la **capa límite** terrestre se conoce como capa superficial, representando la zona de mayor actividad meteorológica. * La **capa límite** planetaria separa la **parte** inferior de la troposfera (donde ocurre la mayor parte de la actividad meteorológica) de la estratosfera. La **capa límite** planetaria - la **parte** más baja de la atmósfera que llega hasta la tropopausa. La **capa límite** terrestre **constituye** la parte más baja de la troposfera. Es la zona de mayor actividad meteorológica.

Figure 7. *Capa límite* and meronymic KPs

Table 1. Definitions of *boundary layer* and *capa límite*

Boundary layer	Capa límite
In physics and fluid mechanics, a boundary layer is the layer of fluid in the immediate vicinity of a bounding surface where the effects of viscosity are significant.	En mecánica de fluidos, la capa límite o capa fronteriza de un fluido es la zona donde el movimiento de este es perturbado por la presencia de un sólido con el que está en contacto.

Both terms thus seem to point to the same concept in both systems. The concordances in Figures 4–7 also provide new hints to continue the analysis. Two other hyponyms of BOUNDARY LAYER can be found in *ocean surface boundary layer* and *capa límite terrestre*, which also shows a meronym in *capa superficial*. It can be surmised that *capa límite atmosférica*, *atmospheric boundary layer* and ABL are term variants of *capa límite planetaria*, *planetary boundary layer* and PBL, since all are described as part of the troposphere and can be confirmed by searching for synonymic patterns (e.g., *also known as*). In fact, in the corpus, even the hypernym *boundary layer* is commonly used as a variant: *This region is known as the atmospheric boundary layer of our planet, or the planetary boundary layer or simply “boundary layer”*.

At this stage, correspondences can be verified or new (translation) equivalents can be found by querying parallel corpora, such as *capa limitrofe del planeta* for *capa límite planetaria* or *capa atmosférica divisoria* for *capa límite atmosférica*. As a final step, the behavior of each identified variant can be compared. For instance, *capa limitrofe* is less frequent than *capa límite*; *capa atmosférica divisoria* is only the result of translation. In formal terms, *capa límite planetaria* is closer to *planetary boundary layer* than *atmospheric boundary layer*, but their degree of equivalence should not be different unless they show clear usage differences. After comparing their contexts, collocations and sources, the only difference is frequency (*atmospheric boundary layer* shows more hits). In contrast, total equivalence of acronym variants (PBL and ABL) depends on their respective full forms.

The previous example shows how total equivalence can be established for a candidate pair. At the other end of the spectrum, non-equivalence cases are those nodes of the concept system with no correspondence. However, in case conceptual analysis was not thorough enough, non-matched nodes should still be further re-searched in comparable and parallel corpora.

For example, in comparable corpora, correspondences for MWTs (Cabezas-García and León-Araúz 2021) can be found by looking for the equivalent(s) of the head term, sorting the results to generate a list of MWTs in the target language, and searching for possible coinciding elements among all modifiers. Another possibility is to look for the possible equivalents of heads and modifiers in different orders and

with optional elements in between or looking for terms to which the MWT is semantically related together with KPs. Finally, when looking for the term in parallel corpora, new translation variants may even emerge. Table 2 illustrates these steps.

Table 2. Looking for correspondences in comparable and parallel corpora

Source term	Query	Equivalent
Long range transboundary air pollution	Query In comparable corpus Head term: <i>contaminación / polución</i>	Contaminación atmsosférica transfronteriza a larga distancia
Long range transboundary air pollution	Query In comparable corpus Head and modifiers: <i>contaminación / polución (...) rango / distancia / alcance</i>	Contaminación atmsosférica transfronteriza a larga distancia, contaminación atmsosférica transfronteriza a largas distancias, contaminación atmsosférica transfronteriza a gran distancia, contaminación atmsosférica transfronteriza de largo alcance
Low level ozone	Query In comparable corpus Causes of source term and KPs: <i>compuesto orgánico volátil (...) crear / producir / generar</i>	Ozono en el nivel del suelo Ozono en la baja atmósfera Ozono troposférico
Long range transboundary air pollution	Query In parallel corpus	Contaminación atmsosférica transfronteriza de gran distancia, contaminación atmsosférica transfronteriza de larga distancia, polución atmsosférica transfronteriza de largo alcance, contaminación atmsosférica Internacional de largo alcance

If KP-based searches retrieve different results in each language, partial equivalence might apply and a comparison of conceptual structures can be performed. If neither comparable nor parallel corpora retrieve any results, then a translation equivalent could be created. If the concept exists in both language cultures but only one has named it, new MWTs can be created by compounding, provided that smaller units in the MWT are documented in the target language. If the concept does not exist in one of the language cultures, terminographers could then propose a translated term to make the concept accessible to other cultures. This means that they must decide whether to emphasize similarities (e.g., functional/cultural adaptation) or differences (e.g., description) (Arntz 1993).

At first, non-equivalence may seem the greatest challenge when building a terminographic resource. However, the problem of non-equivalence has clear-cut

solutions (e.g., loanwords, calques, descriptions, neonyms), depending on the contact between the source and target language cultures as well as the communicative situation or usual conventions in the target domain. The real challenge lies in partial equivalence (i.e. inclusion and overlapping), because it requires extensive research and determines resource design as well as the choice of terms to be matched. In fact, classical approaches to Terminology regard partial equivalence more as a problem to avoid than as a way to account for cultural differences (Van Campenhoudt 2001).

Van Campenhoudt (2001) criticizes the classical triangular model in Terminology on the grounds that it does not account for the referential and contextual aspects underlying equivalence, such as the fact that language cultures classify reality according to different categorization principles.

In this sense, Janssen and Van Campenhoudt (2005) illustrate how, in multilingual terminography, anisomorphism, together with a concept-oriented approach, determine the creation of different entries that would be merged into one entry in a monolingual setting. In the *Dictionnaire hydrographique* (O.H.I. 1994–1998), they find that the concept ENTRÉE has two entries only because in English and Spanish two different senses are split into two terms (*entrance* and *mouth* / *entrada* and *desembocadura*) where the only difference is whether ENTRÉE refers to an artificial entity (channel, port, dock), or a natural entity (ocean, bay).

This is also commonplace in multilingual thesauri and other knowledge organization systems. According to Riesthuis (2005), the most common view is that all language versions of a multilingual thesaurus should be identical and symmetrical. As a consequence, it often happens that cross-language equivalences are forced, where they do not exist, and questionable relational structures occur.

In a concept-oriented terminographic resource, this kind of partial equivalence (i.e., inclusion) can be addressed by: (i) considering the French term a hypernym of English and Spanish terms, which would not have a French equivalent (Figure 8); or (ii) considering the French term partial equivalents of all four terms (Figure 9), which Van Campenhoudt calls *hyperonomase* (2001).

Strictly speaking, the first option (Figure 8) matches the original take on conceptual equivalence, but in that case, three entries in the resource are left without equivalents. It thus seems fair that in cases of anisomorphism, the languages that show more separate senses should prevail over the other(s) (Figure 9). However, establishing equivalence by *hyperonomase* may lead to a proliferation of meanings (Janssen 2003) and also has disadvantages for translation purposes. For instance, if an English text were to make the distinction between *entrance* and *mouth* explicit, translators would only have access to one term in French (*entrée*). They would need to interpret the content of the resource in order to create new terms that lexicalize

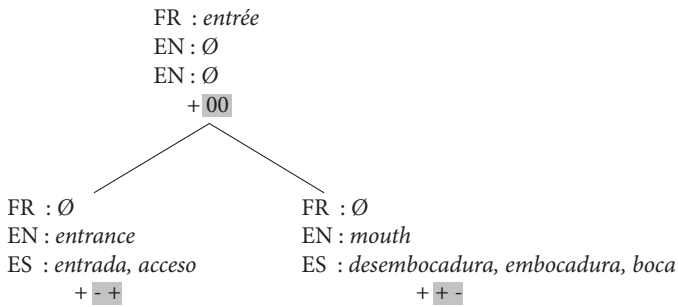


Figure 8. *Entrée* as a hypernym (Van Campenhoudt 2001: 15)

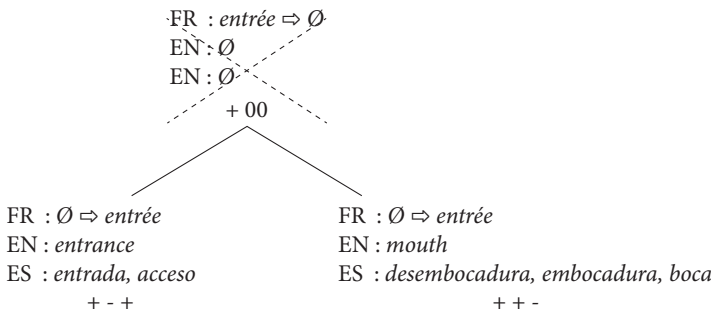


Figure 9. *Entrée* as a partial equivalent (Van Campenhoudt 2001, 16)

the differentiating features in the source language, such as *entrée d'une baie* and *entrée d'une installation portuaire* (Van Campenhoudt 2001, 16). Terminographers could thus anticipate and include such translation equivalents.

Including this kind of equivalents compromises the reversibility of entries, because in many cases a translation equivalent does not refer to an authentic concept in the target culture (Sager 1990, 139). However, adding a note related to usage and directionality could help to identify the gap and clarify the terminographer's solution. In any case, if these terms were not included, or if hyperonymase was not applied, reversibility would still be compromised because of empty nodes.

4.2 Representing equivalence

Depending on how equivalence is established, the resulting terminographic resource can vary in regard to the choice of relational model and associated data categories in both concept and term entries. For instance, if equivalence is established based on a concept-oriented analysis, conceptual networks can be implemented, and the different terms in all languages designating the same concept would be stored under the same record. Based on the scope of equivalence, the number and nature of equivalents/correspondences included might change, and labels for the degrees of equivalence could possibly be added. Likewise, the potential users of terminographic resources also determine how equivalence is understood by terminographers and how concepts, terms, variants and equivalents are described (e.g., adding diverse usage-related notes, metalinguistic information, etc.).

As stated by Rogers (2007a), the varying degrees of equivalence can be reflected in a resource through a combination of definitions in each language combined with an additional remark to highlight differences. In this sense, COTSOES (2003) recommends the indication of the degree of equivalence by means of a note or symbols:

1. = concepts with the same intension
2. ~ concepts with roughly the same intension
3. > the concept has a smaller intension than the other
4. < the concept has a larger intension than the other

In the same line, Šarčević (1989) proposes the use of $\approx$ for near-equivalence, $\pm$ for partial equivalence and $\neq$ for non-equivalence. However, she also argues in favor of including additional guidelines to assist translators when there is conceptual incongruence (Šarčević 1988), since merely including functional equivalents (inherently incongruent) with no other explanations is not sufficient. She illustrates this with the example of *hypothèque* and *mortgage* in Termium Plus (2020). Although the resource warns that the two terms are incongruent, no other guidelines are provided to assist translators in knowing that *mortgage* can be applied to both personal and real property and *hypothèque* only to the latter. According to Šarčević (1989), legal entries should include, apart from definitions, essential differences and similarities in scope of application, structure and legal effect as well as guidelines to usage and/or concrete proposals on how to compensate for the incongruences of functional equivalents.

The approach taken by the EuroTermBank is similar, as the degree of equivalence between terms of different languages is indicated with a set of data categories such as false friend, directionality, and transfer comment (Vasiljevs and Rirdance 2008). Sager (1990, 224) also recommends indicating degrees of equivalence in a

note as well as the contexts in which the use is appropriate, since equivalents may not entirely match the definition of the first language. In fact, it has been questioned whether a term in a multilingual term bank can be matched with a definition, rather than the equivalent term, in another language (Gaballo 2010). For this reason, in multilingual terminology work, definitions in all languages involved are usually provided, as is the case in major resources such as IATE (Inter-Active Terminology for Europe) or Termium Plus, where definitions are accompanied by an authoritative source in each language culture.

In this same line, Sager (1990, 152) advises adding a term source which matches the source of the entry term so that an equivalence of usage can be demonstrated. If there is no original source available in the foreign language, then it is useful to indicate whether (i) the term has been coined by the translator or terminologist – in which case it can be considered a valid equivalent only with respect to the source of the entry term; (ii) whether it is a translation equivalent only documented in a dictionary but not found in text; or (iii) whether it is a neologism in general use as translation equivalent.

Following Sager's recommendations, IATE and Termium Plus also include notes, contexts, and labels. For instance, Termium Plus (2020) provides valuable information regarding equivalence in the field *POLLUTION*, as found in the *POLLUTION* entry, where conceptual incongruence is addressed at the intralingual level: (1) *Pollution is very often used in the more general sense of "environmental pollution".* (2) *In water pollution, the term contamination is often used as a synonym (...) However, a distinction should be made between "pollution" and "contamination": the latter implies a health danger (particularly to humans).* Nevertheless, as reported by Šarčević (1989), incongruence is not always addressed in this resource, as in the *GROIN* entry, where a myriad of partial equivalents are included (*groyne, jetty, spur dike, spurdike, spur, wing dam, bankhead, transverse dike*) with no further explanations, probably because of the effort to reconcile asymmetries in English, French and Spanish.

Other labels that users can use to assess equivalence are acceptability rating (correct, avoid, unofficial), status (standardized, officially approved) or semantic parameters (generic, specific). In IATE, similar labels are evaluation (preferred, admitted, deprecated) and reliability (not verified, minimum reliability, reliable, very reliable). In the field, *Language usage*, information on conceptual incongruence is provided, as in the *OCULAR ULCER* entry: *The term ocular ulcer is seldom used when referring to animals, although it occurs in CELEX:32006R1950/EN where it is used for horses. When referring to humans, this term is used to denote a broader concept, encompassing both corneal and eyelid ulcers.*

Criticizing the rigidity of terminological data banks back in the 90s, Sager (1990, 171) states that the inability to represent the relationship “can be translated in language X by” as a many:many relationship resulted in the following: (i) the need to declare each concept as having only one equivalent concept in each foreign language without inferring that the concepts match; (ii) entering foreign language equivalents only for matching concepts; (iii) storing target language equivalents with a limited amount of associated data as part of a monolingual record in the source language database; (iv) implementations of network-type data models preclude the dynamic creation of new record types and new relationships.

Advocating a many:many relationship implies widening the scope of equivalence from a descriptive lexicon-driven perspective, including the simultaneous representation of genuine and translation equivalents and other term variants as well as the relations holding among them. For translation purposes, referring a term variant to its relevant concept only covers the first half of the task, namely understanding, or analyzing the text. In the text production half of the translation process, it is necessary to be able to match variant with variant and this requires two conditions to be fulfilled (Sager 1994b, 56): (i) identify the contextual situation in the source language which requires a particular type of variant and (ii) identify the rhetorical or stylistic requirements for such variants in the target language to occur in corresponding situations. In this line, Rogers (1997) underlines that translators are frequently obliged to make decisions not only regarding synonymy within the source text and target text, but also regarding the cross-linguistic relations between these synonyms.

Other voices, such as Sandrini (1996), claim that the objective of terminology is not equivalence in the sense of complete interchangeability in text –a target which could only be reached in a few cases– it is rather conceptual correspondence. However, describing the interchangeability of terms can also be among terminographic aims. After all, translation term pairs are hardly ever symmetric and interchangeability can be highly dynamic (León-Araúz and Faber 2014). In the same way that degrees of equivalence can be specified with regards to a concept, the differences and similarities among terms can also be accounted for in a resource. In fact, in her legal terminology resource MuLex, Peruzzo (2012) specifies a synonymy degree between term variants and main entry terms with the symbols ~ and < and includes what she calls *Notes on terminological variation* highlighting conceptual asymmetries, vagueness, regional differences, etc. Pimentel (2012, 70) agrees that a suitable resource for translators would document the concepts of the specialized domain but also the linguistic behavior of terms. In EcoLexicon (Faber, León-Araúz and Reimerink 2016, Faber in this volume), a multilingual terminological knowledge base on environmental science, a recent proposal has been designed for the

representation of interlingual and intralingual equivalence at both concept and term levels (León-Araúz, Cabezas-García and Reimerink 2020). However, contrary to what the traditional models would have advocated for, most of the data categories that address equivalence issues are at the term level.

At the concept level, equivalence is rather addressed as a design issue with regards to the resource's macrostructure and the choice of equivalent terms. It has generally followed the hyperonymase approach. For instance, two entries were created for the concept CANAL because in English there is a distinction between a natural one (CHANNEL) and an artificial one (CANAL). In contrast, for more specific culture/territory-bound concepts, the approach is the opposite. No different entries are created for the Spanish term *pororoca*, which is a specific type of TIDAL BORE in the Amazon River. However, this information is included in a term-level note. The conceptual networks are fully reversible only in English and Spanish, which involved naming the concepts in a somewhat artificial way in order to overcome polysemy issues (i.e., the name of the main entry term followed by a contextual disambiguating feature instead of the usual numbers).

At the term level, different data categories provide information in order to discriminate among term variants and to reflect the degrees of concept and term equivalence. Figures 10–12 are examples of how the new fields would describe at the term level, *ozono troposférico*, *ozono a nivel del suelo* and *ozono superficial*, all Spanish variants of *tropospheric ozone*, also known as *ground-level ozone*, *surface ozone*, and *low level ozone*. For instance, Use_context includes any information about the nuances that a particular variant may have in comparable corpora, whereas Use_translation context is filled when clear patterns are found regarding the asymmetries of equivalence in parallel corpora.

For example, although *ozono troposférico* is clearly the most frequent Spanish variant designating TROPOSPHERIC OZONE, in comparable corpora, *ozono a nivel del suelo* and *ozono superficial* seem to be preferred when the term is related to human health issues. In parallel corpora, although *ozono troposférico* is usually translated by *tropospheric ozone* or *ground-level ozone*, *ozono a nivel del suelo* and *ozono superficial* clearly prefer *ground-level ozone*, even though it is exactly the same concept. Consequently, the field Use_translation context establishes interlinguistic variation preferences whereas the field Use_context serves the same purpose for intralinguistic variation.

Term information	
Term	Ozono troposférico
Term type	Main term
Formation device	
Source	IATE, TermiumPlus, EcoLexicon corpus, EurLex...
Use_geographical	
Use_status	Standardized
Use_register	Formal
Use_context	
Use_translation context	Usually translated by <i>tropospheric ozone</i> , although <i>ground-level ozone</i> can also be found in institutional settings (e.g. European Parliament, European Commission, UN).
Notes	
View concordances	

Term information	
Term	Ozono a nivel del suelo
Term type	Variant
Formation device	Calque
Source	EcoLexicon corpus, EurLex, OPUS
Use_geographical	Especially used in Latin America
Use_status	Admitted
Use_register	Formal semi-specialized
Use_context	This term can be used instead of <i>ozono troposférico</i> when focusing on its effects on human health, as "a nivel del suelo" gives a better perspective on the proximity of the chemical compound to the soil.
Use_translation context	Usually translated as <i>ground-level ozone</i> .
Notes	
View concordances	

Term information	
Term	Ozono superficial
Term type	Variant
Formation device	Calque
Source	EcoLexicon corpus, EurLex, OPUS
Use_geographical	
Use_status	Admitted
Use_register	Formal semi-specialized
Use_context	This term can be used instead of <i>ozono troposférico</i> when focusing on its effects on human health, as "superficial" gives a better perspective on the proximity of the chemical compound to the surface.
Use_translation context	Usually translated as <i>ground-level ozone</i> or <i>surface ozone</i> .
Notes	
View concordances	

Figure 10–12. Term entries for *ozono troposférico*, *ozono a nivel de suelo* and *ozono superficial*

5. Conclusions

Despite the traditional lack of interest in Terminology, equivalence is a complex notion that can be understood from different perspectives and with different consequences for the development of terminographic products.

Ideally, for equivalence to be well represented in descriptive terminographic resources, all different term equivalents/correspondences and variants should be accounted for and differentiated, which would entail determining the relation between intralingual variants and the concept, and between interlingual and intralingual correspondences. In this way, having a set of different terms pointing to the same concept would cover different types of equivalence; and including notes, contexts, and data categories related to usage would be useful for the user to ascertain equivalence degrees in different possible contexts and make the right (informed) choices.

However, further research should be devoted to methodological issues in order to make resources as consistent and useful as possible. Even though the design of resources is constrained by many factors (e.g., users, specialized domain, languages), a more systematic approach to establishing and representing equivalence is still needed in Terminology.

Funding

This research was carried out as part of the projects PID2020-118369GB-I00, Transversal integration of culture into an environmental terminological knowledge base (TRANSCULTURE) and A-HUM-600-UGR20, Culture as a transversal module in an environmental terminological knowledge base (CULTURAMA) funded by the Spanish Ministry of Science and Innovation and the Regional Government of Andalusia.

References

A

- Adamska-Sałaciak, Arleta. 2010. "Examining Equivalence." *International Journal of Lexicography* 23 (4): 387–409. <https://doi.org/10.1093/ijl/ecq024>
- Adamska-Sałaciak, Arleta. 2011. "Between Designer Drugs and Afterburners: A Lexicographic-semantic Study of Equivalence." *Lexikos* 21: 1–22.
- Adamska-Sałaciak, Arleta. 2013. "Equivalence, Synonymy, and Sameness of Meaning in a Bilingual Dictionary." *International Journal of Lexicography* 26 (3): 329–345. <https://doi.org/10.1093/ijl/ect016>
- Adamzik, Kirsten. 2000. *Textsorten*. Tübingen: Stauffenburg.
- Adelstein, Andreína. 2007. "Unidad léxica y significado especializado: modelo de representación a partir del nombre relacional madre." PhD diss., Universitat Pompeu Fabra.
- Adelstein, Andreína, and Maria Teresa Cabré. 2002. "The Specificity of Units with Specialized Meaning: Polysemy as Explanatory Factor." *D.E.L.T.A.* 18: 1–25. <https://doi.org/10.1590/S0102-44502002000300003>
- Agbago, Akakpo, and Caroline Barrière. 2005. "Corpus Construction for Terminology." In *Proceedings from the Corpus Linguistics Conference Series* 1(1), edited by Anna Mauranen, Mike Scott, and Alison Wray. Birmingham: University of Birmingham.
- Aguado-de-Cea, Guadalupe, Asunción Gómez-Pérez, Elena Montiel-Ponsoda, and Mari Carmen Suárez-Figueroa. 2008. "Natural Language-based Approach for Helping in the Reuse of Ontology Design Patterns." In *Proceedings of the 16th International Conference on Knowledge Engineering (EKAW08)*, 32–47. Catania, Italy.
- Aguado-de-Cea, Guadalupe, Inmaculada Álvarez de Mon, and Elena Montiel-Ponsoda. 2009. "From Linguistic Patterns to Ontology Structures." In *Proceedings of the 8th International Conference on Terminology and Artificial Intelligence*, edited by Marie-Claude L'Homme, and Sylvie Szulman. Toulouse, France.
- Ahlthorp, Magnus, Maria Skeppstedt, Shiho Kitajima, Aron Henriksson, Rafal Rzepka, and Kenji Araki. 2016. "Expansion of Medical Vocabularies Using Distributional Semantics on Japanese Patient Blogs." *Journal of Biomedical Semantics* 7: 58. <https://doi.org/10.1186/s13326-016-0093-x>
- Ahlthorp, Magnus, Maria Skeppstedt, Shiho Kitajima, Rafal Rzepka, and Kenji Araki. 2014. "Medical Vocabulary Mining Using Distributional Semantics on Japanese Patient Blogs." In *Proceedings of the 6th International Symposium on Semantic Mining in Biomedicine (SMBM 2014)*, edited by Olivier Bodenreider, José Luis Oliveira, and Fabio Rinaldi, 69–73. Aveiro, Portugal. <https://doi.org/10.5167/UZH-98982>
- Ahmad, Khurshid, and Heather Fulford. 1992. "Knowledge Processing: 4. Semantic Relations and their Use in Elaborating Terminology." *Computing Sciences Report CS-92-07*. Guildford: University of Surrey.

- Ahmad, Khurshid, and Margaret Rogers. 2001. "Corpus Linguistics and Terminology Extraction." In *Handbook of Terminology Management* Vol. 2, edited by Sue Ellen Wright, and Gerhard Budin, 725–760. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/z.htm2.28ahm>
- Aito, Emmanuel. 2000. "Terminologie, dénomination et langues minoritaires face à la modernité : vers une interrogation soucieuse du social." *Terminologies nouvelles* 21: 46–51.
- Alarcón, Rodrigo, Carme Bach, and Gerardo Sierra. 2007. "Extracción de contextos definitorios en corpus especializados: hacia la elaboración de una herramienta de ayuda terminográfica." *Revista Española de Lingüística (RSEL)* 37: 247–277.
- Alexandru, Cristina, and François Gaudin. 2006. "Les contextes : à la source du terme." In *Mots, termes et contextes. Actes des septièmes journées scientifiques du réseau des chercheurs Lexicologie Terminologie Traduction*, edited by Daniel Blampain, Philippe Thoiron, and Marc Van Campenhoudt, 59–67. Paris: Editions des archives contemporaines.
- Alexiev, Boyan. 2005. "Contrastive Aspects of Terminological Metaphor." PhD diss., University of Sofia.
- Almeida, Mauricio, Renato Souza, and Fred Fonseca. 2011. "Semantics in the Semantic Web: A Critical Evaluation." *Knowledge Organization* 8 (3): 187–203.
<https://doi.org/10.5771/0943-7444-2011-3-187>
- Alonso Campo, Araceli, and Irene Renau Araque. 2013. "Corpus Pattern Analysis in Determining Specialised Uses of Verbal Lexical Units." *Terminalia* 7: 26–33.
- Alrashidi, Badriyah. 2016. "Un problème de terminologie médicale Au Koweït : la vulgarisation autour du diabète." PhD diss., Université de Rouen.
- Alrashidi, Badriyah, Valérie Delavigne, and François Gaudin. 2018. "La référencement des termes : perspectives socioterminologiques." In *Terminologie et Ontologie. Théories et Applications*, edited by Christophe Roche, 35–50. Chambéry: Presses Universitaires Savoie Mont-Blanc.
- Amjadian, Ehsan, Diana Inkpen, T. Sima Paribakht, and Farahnaz Faez. 2018. "Distributed Specificity for Automatic Terminology Extraction." *Terminology* 24 (1): 23–40.
<https://doi.org/10.1075/term.00012.amj>
- Ananiadou, Sophia. 1994. "A Methodology for Automatic Term Recognition." In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 1994)*, 1034–1038. Stroudsburg, PA: Association for Computational Linguistics.
<https://doi.org/10.3115/991250.991317>
- Anderson, John R. 1983. *The Architecture of Cognition*. Hillsdale, NJ: Erlbaum.
- Andreasen, Troels, Per Anker Jensen, J. Fischer Nilsson, Patrizia Paggio, Bolette Sandford Pedersen, and Hanne Erdman Thomsen. 2004. "Content-Based Text Querying with Ontological Descriptors." *Data & Knowledge Engineering* 48: 199–219.
[https://doi.org/10.1016/S0169-023X\(03\)00105-8](https://doi.org/10.1016/S0169-023X(03)00105-8)
- Anis, Jacques (ed.). 2009. *Communication et langue française*. Paris: Hermès.
- Anthony, Kelsy C., and Daniel Pomeranz Krummel. 2013. "Spliceosome." In *Encyclopedia of Life Science, eLS*. Hoboken NJ: John Wiley and Sons Ltd.
<https://doi.org/10.1002/9780470015902.a0005323.pub2>
- Anthony, Lawrence. 2018. "Visualisation in Corpus-based Discourse Studies." In *Corpus Approaches to Discourse: A Critical Review*, edited by Charlotte Taylor, and Anna Marchi, 197–224. London: Routledge. <https://doi.org/10.4324/9781315179346-10>
- Antia, Bassey E. 2000. *Terminology and Language Planning: An Alternative Framework of Practice and Discourse*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.2>

- Antia, Bassey E. 2001. "Metadiscourse in Terminology: Thesis, Antithesis, Synthesis." *Terminology Science and Research* 12 (1–2): 65–84.
- Antia, Bassey E. 2002. "Wüster and (Applied) Linguistic Research." Review of *Terminologie und Wissensordnung*, edited by Heribert Picht, and Klaus Dirk Schmidt, *LPS and Professional Communication* 2 (1): 102–111.
- Apresjan, Juri. 1974. "Regular Polysemy." *Linguistics* 142: 5–32.
<https://doi.org/10.1515/ling.1974.12.142.5>
- Arndt, Suzanne, and Dieter Schnäpp. 2020. "Harbsafe-162 – A Domain-Specific Data Set for the Intrinsic Evaluation of Semantic Representations for Terminological Data." arXiv: 2005.14576v1 [cs.IR].
- Arntz, Reiner. 1985. "Review of Rondeau *Introduction à la terminologie*, 2^e édition." *Knowledge Organisation* 12 (1): 44–45. <https://doi.org/10.5771/0943-7444-1985-1-44>
- Arntz, Reiner. 1993. "Terminological Equivalence and Translation." In *Terminology. Applications in Interdisciplinary Communication*, edited by Helmi B. Sonneveld, and Kurt L. Loening, 5–19. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.70.02arn>
- Arntz, Reiner, and Heribert Picht. 1989a. *Einführung in die Terminologearbeit*. Hildesheim, Zürich/New York: Georg Olms Verlag.
- Arntz, Reiner, and Heribert Picht. 1989b. *Introducción a la terminología*. Translated by Amelia de Irazazábal et al. Madrid: Ediciones Pirámide.
- Asim, Muhammad Nabeel, Muhammad Wasim, Muhammad U. G. Khan, Waqar Mahmood, and Hafiza Mahnoor Abbasi. 2018. "A Survey of Ontology Learning Techniques and Applications." *Database* Vol. 2018, bay101. <https://doi.org/10.1093/database/bay101>
- Assal, Allal. 1992. *Vocabulaire des biotechnologies: une approche descriptive*. Rouen: Université de Rouen.
- Atkins, B. T. Sue. 2008. "Theoretical Lexicography and its Relation to Dictionary-Making." In *Practical Lexicography, A Reader*, edited by Thierry Fontenelle, 31–50. Oxford/New York: Oxford University Press.
- Atkins, B. T. Sue, and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.
- Auger, Alain, and Caroline Barrière. 2008. "Pattern-based Approaches to Semantic Relation Extraction: A State-of-the-Art." *Terminology* 14 (1): 1–19.
<https://doi.org/10.1075/term.14.1.02aug>
- Auger, Pierre. 1974. "Observation de la synonymie dans la terminologie minière." In *La normalisation linguistique. Actes du colloque international de terminologie*, 25–33. Québec: Office de la langue française.
- Auger, Pierre. 1999. "L'implantation des officialismes halieutiques au Québec: essai de terminométrie." *Langues et Sociétés* 37. Québec: Office de la langue française.
- Auger, Pierre. 2006. "Le phénomène de l'anglicisation de la langue forestière au Québec: essai de socioterminologie diachronique." In *Aspects diachroniques du vocabulaire*, edited by Danielle Candel, and François Gaudin, 137–156. Rouen: Publications des Universités de Rouen et du Havre.
- Auger, Pierre, Patrick Drouin, and Marie-Claude L'Homme. 1991. "Automatisation des procédures de travail en terminographie." *Meta* 36 (1): 121–127. <https://doi.org/10.7202/001921ar>
- Auger, Pierre, and Louis-Jean Rousseau. 1978. *Méthodologie de la recherche terminologique*. Québec: Office de la langue française.
- Aussenac-Gilles, Nathalie, Brigitte Biébow, and Sylvie Szulman. 2000a. "Modélisation du domaine par une méthode fondée sur l'analyse du corpus." In *Actes de la Conférence IC'2000, Journées Francophones d'Ingénierie de la Connaissance*, 93–104. Toulouse: IRIT.

- Aussenac-Gilles, Nathalie, Brigitte Biébow, and Sylvie Szulman. 2000b. "Revisiting Ontology Design: A Method Based on Corpus Analysis." In *Proceedings of the European Knowledge Acquisition Workshop, EKAW 2000: Knowledge Engineering and Knowledge Management Methods, Models, and Tools*, edited by Rose Dieng, and Olivier Corby, 172–188. Juan-les-Pins, France. https://doi.org/10.1007/3-540-39967-4_13
- Aussenac-Gilles, Nathalie, Didier Bourigault, and Anne Condamines. 1995. "How can Knowledge Acquisition Benefit from Terminology?" In *Proceedings of the 9th Knowledge Acquisition Workshop, Advances in Knowledge Acquisition*, edited by Brian R. Gaines, and Mark Musen, 1–19. Banff, Alberta.
- Aussenac-Gilles, Nathalie, and Anne Condamines. 2012. "Variation and Semantic Relation Interpretation: Linguistic and Processing Issues." In *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE 2012)*, edited by Guadalupe Aguado-de-Cea, Mari Carmen Suárez-Figueroa, Raúl García-Castro, and Elena Montiel-Ponsoda, 106–122. Madrid, Spain.
- Aussenac-Gilles, Nathalie, Sylvie Després, and Sylvie Szulman. 2008. "The TERMINAE Method and Platform for Ontology Engineering from Texts." In *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*, edited by Paul Buitelaar, and Philipp Cimiano, 199–223. Amsterdam: IOS Press.
- Aussenac-Gilles, Nathalie, and Marie-Paule Jacques. 2008. "Designing and evaluating patterns for relation acquisition from texts with CAMÉLÉON." *Terminology* 14 (1): 45–73. <https://doi.org/10.1075/term.14.1.04aus>
- Azoulay, Daphnée. 2017. "Frame-Based Knowledge Representation Using Large Specialized Corpora." In *Proceedings of the AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding* (Technical report SS-17-02), edited by Jerome Feldman, and Luc Steels, 119–126. Palo Alto: AAAI Press.
- B**
- Baayen, Harald R. 2001. *Word Frequency Distributions*. Dordrecht: Kluwer. <https://doi.org/10.1007/978-94-010-0844-0>
- Bachimont, Bruno. 2000. "Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en Ingénierie des connaissances." In *Ingénierie des connaissances, Evolutions récentes et nouveaux défis*, edited by Jean Charlet, Manuel Zacklad, Didier Bourigault, and Gilles Kassel, 305–323. Paris: Eyrolles.
- Baker, Mona. 1992. *In Other Words: A Coursebook on Translation*. London: Routledge.
- Baker, Mona. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 7 (2): 223–243. <https://doi.org/10.1075/target.7.2.03bak>
- Banks, David. 2017. *The Birth of the Academic Article, Le Journal des Sçavans and the Philosophical Transactions, 1665–1700*. Sheffield: Equinox Publishing Ltd.
- Barabási, Albert-László, and Reka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286: 509–512. <https://doi.org/10.1126/science.286.5439.509>
- Baraké, Rima. 2016. "Le terme et ses changements linguistiques: entre conceptualisation et transformations vulgarisatrices." In *Repères Dorif n° 10. Le terme un produit social?* DoRif Università (<https://www.dorif.it/reperes/rima-barake-le-terme-et-ses-changements-linguistiques-en-tre-conceptualisation-et-transformations-vulgarisatrices/>). Accessed 7 December 2021.
- Baroni, Marco. 2016. "Grounding Distributional Semantics in the Visual World." *Language and Linguistics Compass* 10 (1): 3–13. <https://doi.org/10.1111/lnc3.12170>

- Baroni, Marco, Georgiana Dinu, Germán Kruszewski. 2014. "Don't Count, Predict! A Systematic Comparison of Context-counting vs. Context-predicting Semantic Vectors." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL)*, Vol. 1, edited by Kristina Toutanova, and Hua Wu, 238–247. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/P14-1023>
- Baroni, Marco, and Alessandro Lenci. 2010. "Distributional Memory: A General Framework for Corpus-based Semantics." *Computational Linguistics* 36 (4): 673–721. https://doi.org/10.1162/coli_a_00016
- Barrière, Caroline. 2001. "Investigating the Causal Relation in Informative Texts." *Terminology* 7 (2): 135–154. <https://doi.org/10.1075/term.7.2.02bar>
- Barrière, Caroline. 2004. "Knowledge-Rich Contexts Discovery." In *Proceedings of the Seventeenth Canadian Conference on Artificial Intelligence LNAI 360*, London, ON, May 17–19, 2004, edited by Ahmed Y. Tawfik, and Scott D. Goodwin, 187–201. Berlin: Springer. https://doi.org/10.1007/978-3-540-24840-8_14
- Barrière, Caroline. 2006. "Semi-automatic Corpus Construction from Informative Texts." In *Lexicography, Terminology and Translation: Text-based Studies in Honour of Ingrid Meyer*, edited by Lynne Bowker, 81–92. Ottawa: University of Ottawa Press. <https://doi.org/10.2307/j.ctt1ckpgs3.9>
- Barsalou, Lawrence W. 1992. "Frames, Concepts, and Conceptual Fields." In *Frames, Fields, and Contrasts*, edited by Adrienne Lehrer, and Eva Feder Kittay, 21–74. Hillsdale, NJ: Erlbaum.
- Barsalou, Lawrence W. 2003. "Situated Simulation in the Human Conceptual System." *Language and Cognitive Processes* 18 (5–6): 513–562. <https://doi.org/10.1080/016909603440000026>
- Barsalou, Lawrence W. 2017. "Cognitively Plausible Theories of Concept Composition." In *Compositionality and Concepts in Linguistics and Psychology*, edited by James A. Hampton, and Yoad Winter, 9–30. Berlin: Springer. https://doi.org/10.1007/978-3-319-45977-6_2
- Barsalou, Lawrence W., Paula M. Niedenthal, Aron K. Barbey, and Jennifer A. Ruppert. 2003. "Social Embodiment." In *The Psychology of Learning and Motivation: Advances in Research and Theory*, edited by Brian H. Ross, 43–92. Vol. 43. Amsterdam: Elsevier Science.
- Baudet, Jean-Claude. 1991. "Éditologie et sociolinguistique." *Cahiers de linguistique sociale. Terminologie et sociolinguistique* 18: 81–100.
- Baumann, Klaus-Dieter. 2005. "Das komplexe Normensystem der Fachkommunikation." *Fachsprache* 27: 32–47.
- Beeson, Diane, Marcy Darnovsky, and Abby Lippman. 2015. "What's in a name? Variations in Terminology of Third-party Reproduction." *Reproductive Biomedicine Online* 31 (6): 805–814. <https://doi.org/10.1016/j.rbmo.2015.09.004>
- Beger, Anke. 2016. "Different Functions of (Deliberate) Metaphor in Teaching Scientific Concepts." *Metaphorik* 26: 61–86.
- Béjoint, Henri. 1988. "Scientific and Technical Words in General Dictionaries." *International Journal of Lexicography* 1 (4): 354–368. <https://doi.org/10.1093/ijl/1.4.354>
- Béjoint, Henri, and Philippe Thoiron (eds). 2000. *Le sens en terminologie*. Lyon: Presses Universitaires de Lyon.
- Béjoint, Henri, and Philippe Thoiron. 2010. "La terminologie : une question de termes?" *Meta* 55 (1): 105–128. <https://doi.org/10.7202/039605ar>
- Belloncle, Guy. 1984. *La question éducative en Afrique noire*. Paris: Karthala.
- Belloncle, Guy, Peter Easton, Paul Ilboudo, and Papa Sène. 1987. *Alphabétisation et gestion des groupements villageois en Afrique sahélienne*. Paris: Karthala.

- Benali, Abdelkader. 2012. "Les problèmes de la catégorisation textuelle: entre fondements théoriques et fondements structurels." *Synergies Algérie* 17: 35–49.
- Bennett, Gena R. 2010. *Using Corpora in the Language Learning Classroom: Corpus Linguistics for Teachers*. Ann Arbor MI: University of Michigan Press ELT.
<https://doi.org/10.3998/mpub.371534>
- Bergmann, Joerg, and Thomas Luckmann. 1995. "Reconstructive Genres." In *Aspects of Oral Communication*, edited by Uta M. Quasthoff, 289–304. Berlin/New York: Mouton de Gruyter.
<https://doi.org/10.1515/9783110879032.289>
- Berl, Matthew, and Eugene Charniak. 1999. "Finding Parts in very Large Corpora." In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, 57–64. College Park, MA: Association for Computational Linguistics.
- Bernal, Elisenda. 2017. "La variación sobre dos ruedas: *ciclovía* y otros caminos de la bici." In *Antiedad, pansexual, fracking y otras palabras recientes del español de América y España*, edited by Andreína Adelstein, Judit Freixa, Constanza Gerding, Rosa Luna, María Pozzi, and Mercedes Suárez, 122–124. Los Polvorines, Argentina: Ediciones Universidad Nacional de General Sarmiento.
- Bernaras, Amaia, Iñaki Laregoiti, and José Manuel Corera. 1996. "Building and Reusing Ontologies for Electrical Network Applications." In *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI'96)*, 298–302. Chichester, UK.
- Berners-Lee, Tim. 1998. "Q&A with Tim Berners-Lee, Special Report". *businessweek.com* (<https://www.bloomberg.com/news/articles/2007-04-09/q-and-a-with-tim-berners-lee-businessweek-business-news-stock-market-and-financial-advice>). Accessed 3 August 2020.
- Berners-Lee, Tim, Jim Hendler, and Ora Lassila. 2001. "The Semantic Web." *Scientific American* 284 (5): 34–43. <https://doi.org/10.1038/scientificamericano501-34>
- Bernier-Colborne, Gabriel. 2014a. "Analyse distributionnelle de corpus spécialisés pour l'identification de relations lexico-sémantiques." In *Actes de la 21^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2014) Atelier Sémantique Distributionnelle (SemDis 2014)*, edited by Cécile Fabre, Nabil Hathout, Lydia-Mai Ho-Dac, François Morlane-Hondère, Philippe Muller, Franck Sajous, Ludovic Tanguy, and Tim van de Cruys, 238–251. Marseille, France.
- Bernier-Colborne, Gabriel. 2014b. "Identifying Semantic Relations in a Specialized Corpus through Distributional Analysis of a Cooccurrence Tensor." In *Proceedings of the 3rd Joint Conference on Lexical and Computational Semantics (*SEM 2014)*, edited by Johan Bos, Anette Frank, and Roberto Navigli, 57–62. Stroudsburg, PA/ Dublin: Association for Computational Linguistics and Dublin City University. <https://doi.org/10.3115/v1/S14-1007>
- Bernier-Colborne, Gabriel, and Patrick Drouin. 2016. "Evaluation of Distributional Semantic Models: a Holistic Approach." In *Proceedings of the 5th International Workshop on Computational Terminology (Computerm 2016)*, edited by Patrick Drouin, Natalia Grabar, Thierry Hamon, Kyo Kageura, and Koichi Takeuchi, 52–61. Osaka, Japan.
- Bertaccini, Franco, and Alessandra Matteucci. 2005. "L'approche variationniste à la pratique terminologique d'entreprise." *Meta* 50 (4). <https://doi.org/10.7202/019910ar>
- Bertels, Ann, and Dirk Speelman. 2013. "Exploration sémantique visuelle à partir des cooccurrences de deuxième et troisième ordre." In *Actes de la 20^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2013) Atelier Sémantique Distributionnelle (SemDis 2013)*, edited by Cécile Fabre, Nabil Hathout, Philippe Muller, and Tim Van de Cruys, 126–139. Sables d'Olonne, France.

- Bertels, Ann, and Dirk Speelman. 2014a. "Analyse de positionnement multidimensionnel sur le corpus spécialisé TALN." In *Actes de la 21^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2014) Atelier Sémantique Distributionnelle (SemDis 2014)*, edited by Cécile Fabre, Nabil Hathout, Lydia-Mai Ho-Dac, François Morlane-Hondère, Philippe Muller, Franck Sajous, Ludovic Tanguy, and Tim van de Cruys, 252–265. Marseille, France.
- Bertels, Ann, and Dirk Speelman. 2014b. "Clustering for Semantic Purposes: Exploration of Semantic Similarity in a Technical Corpus." *Terminology* 20 (2): 279–303. <https://doi.org/10.1075/term.20.2.07ber>
- Bertels, Ann, and Dirk Speelman. 2016. "Analyse sémantique distributionnelle dans un corpus technique : les enjeux sémantiques dans un jeu de statistiques." In *Actes des Journées internationales d'analyse statistique des données textuelles (JADT 2016)*, edited by Damon Mayaffre, Céline Poudat, Laurent Vanni, Véronique Magri, Peter Follette, Caroline Daire, and Francois Couessurel, 273–283. Nice, France.
- Bestué, Carmen. 2016. "La variación terminológica de los conceptos del derecho de sociedades y sus estrategias de traducción." *Revista de Llengua i Dret*, 65: 18–35.
- Bhatia, Vijay Kumar. 2002. "Applied Genre Analysis: A Multi-Perspective Model." *Ibérica* 4: 3–19.
- Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511519871>
- Biber, Douglas. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.
- Biber, Douglas, and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511814358>
- Biel, Łucja, Agnieszka Biernacka, and Anna Jopek-Bosiacka. 2018. "Collocations of Terms in EU Competition Law: A Corpus Analysis of EU English Collocations." In *Language and Law. The Role of Language and Translation in EU Competition Law*, edited by Silvia Marino, Łucja Biel, Martina Bajčić, and Vilemini Sosoni, 249–274. Cham: Springer. https://doi.org/10.1007/978-3-319-90905-9_14
- Binder, Jeffrey, and Rutvik Desai. 2011. "The Neurobiology of Semantic Memory." *Trends in Cognitive Sciences* 15 (11): 527–536. <https://doi.org/10.1016/j.tics.2011.10.001>
- Binon, Jean, Serge Verlinde, Jan van Dyck, and Ann Bertels. 2000. *Dictionnaire d'apprentissage du français des affaires*. Paris: Didier.
- Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. "Linked Data – The Story So Far." In *Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS)* 5 (3): 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Bodenreider, Oliver, Ronald Cornet, and Daniel J. Vreeman. 2018. "Recent Developments in Clinical Terminologies – SNOMED CT, LOINC, and RxNorm." *Yearbook of Medical Informatics* 27 (1): 129–139. <https://doi.org/10.1055/s-0038-1667077>
- Boisson, Claude. 2001. "Dénomination et 'Vision'" *Cahiers de praxématique* 36: 141–168. <https://doi.org/10.4000/praxematique.348>
- Boleda, Gemma. 2020. "Distributional Semantics and Linguistic Theory." *Annual Review of Linguistics* 6 (1): 213–234. <https://doi.org/10.1146/annurev-linguistics-011619-030303>
- Bomrad, Maria Daniela. 2012. "La variación terminológica del español de la vid y el vino en el espacio geográfico Implicaciones en la traducción especializada." MA Thesis, Faculdade de Letras, Universidade do Porto.
- Bonnadonna, Maria Francesca. 2020. "Using Lexical Functions to Describe Adjectives in Terminography." *Terminology* 26 (1): 7–32. <https://doi.org/10.1075/term.00040.bon>

- Boquera, María. 2005. "Las metáforas en textos de ingeniería civil: Estudio contrastivo inglés-español." PhD diss., University of Valencia.
- Bordet, Geneviève. 2013. "Brouillage des frontières, rencontres des domaines : quelles conséquences pour l'enseignement de la terminologie et de la traduction spécialisée." *ASp. La revue du GERAS* 64: 95–115. <https://doi.org/10.4000/asp.3851>
- Bordet, Geneviève. 2019. "'This Study Explores' ou comment la référence ancre l'abstract dans son contexte énonciatif." *Cahiers de praxématique* 72. <https://doi.org/10.4000/praxematique.5637>
- Borst, Willem Nico. 1997. "Construction of Engineering Ontologies." PhD diss., University of Twente.
- Boschiroli, Victoria de los Ángeles. 2013. "La variación en el discurso jurídico: a propósito de medida cautelar." *Debate Terminológico* 9: 89–91.
- Bosque-Gil, Julia, Elena Montiel-Ponsoda, Jorge Gracia, and Guadalupe Aguado-de-Cea. 2016. "TerminotecaRDF: a Gathering Point for Multilingual Terminologies in Spain." In *Proceedings of the 12th International Conference on Terminology and Knowledge Engineering (TKE 2016)*, 136–146. Copenhagen, Denmark.
- Botov, Dmitriy, Julius Klenin, Andrey Melnikov, Yuri Dmitrin, Ivan Nikolaev, and Mikhail Vinel. 2019. "Mining Labor Market Requirements Using Distributional Semantic Models and Deep Learning." In *Proceedings of the 22nd International Conference on Business Information Systems (BIS 2019)*, edited by Witold Abramowicz, and Rafael Corchuelo, 177–190. New York: Springer. https://doi.org/10.1007/978-3-030-20482-2_15
- Boulanger, Jean-Claude. 1981. "Jean-Claude Corbeil, L'aménagement linguistique du Québec." *Terminogramme* 7–8: 11–12.
- Boulanger, Jean-Claude. 1995a. "Une lecture socioculturelle de la terminologie." *Meta* 40 (2): 216–223.
- Boulanger, Jean-Claude. 1995b. "Présentation: images et parcours de la socioterminologie." *Meta* 40 (2): 194–205. <https://doi.org/10.7202/002117ar>
- Bourigault, Didier, Nathalie Aussenac-Gilles, and Jean Charlet. 2004. "Construction de ressources terminologiques ou ontologiques à partir de textes : un cadre unificateur pour trois études de cas." *Revue d'Intelligence Artificielle (RIA)* 18: 87–110. <https://doi.org/10.3166/ria.18.87-110>
- Bourigault, Didier, and Jean Charlet. 2000. "Ontologie et textes." In *Complément aux Actes de la conférence IC'2000, Journées Francophones d'Ingénierie de la Connaissance*, 7–8. Toulouse: IRIT.
- Bourigault, Didier, and Benoît Habert. 1998. "Evaluation of Terminology Extractors: Principles and Experiments." In *Proceedings from the First International Conference on Language Resources and Evaluation*, edited by Antonio Rubio, Navidad Gallardo, Rosa Castro, and Antonio Tejada, 299–305. Granada: European Language Resources Association.
- Bourigault, Didier, Christian Jacquemin, and Marie-Claude L'Homme (eds). 2001. *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/nlp.2>
- Bourigault, Didier, and Monique Slodzian. 1999. "Pour une terminologie textuelle." *Terminologies nouvelles* 19: 29–32.
- Boutet, Josiane. 1994. *Construire le sens*. Bern/New York: Peter Lang.
- Bouveret, Myriam. 1998. "Approche de la dénomination en langue spécialisée." *Meta* 43 (3): 393–410. <https://doi.org/10.7202/002130ar>
- Bouveret, Myriam, and François Gaudin. 1997. "Du flou dans les catégorisations: le cas de la Bio-informatique." In *Terminologie et Interdisciplinarité: Actes des Troisièmes journées internationales de terminologie*, edited by Caroline De Schatzen, 63–72. Louvain-La-Neuve: Peeters.

- Bowden, Paul R., Peter Halstead, and Tony G. Rose. 1996. "Extracting Conceptual Knowledge from Text Using Explicit Relation Markers." In *Advances in Knowledge Acquisition. Proceedings of the 9th European Knowledge Acquisition Workshop, EKAW'96*, edited by Nigel Shadbolt, Kieron O'Hara, and Guus Schreiber, 147–162. Berlin/Heidelberg: Springer. https://doi.org/10.1007/3-540-61273-4_10
- Bowdle, Brian, and Dedre Gentner. 2005. "The Career of Metaphor." *Psychological Review* 112: 193–216. <https://doi.org/10.1037/0033-295X.112.1.193>
- Bowker, Lynne. 1993. "Multidimensional Classification of Concepts for Terminological Purposes." In *Proceedings of the 4th ASIS SIG/CR Classification Research Workshop, Columbus, OH, October 24, 1993*, edited by Philip J. Smith, Clare Beghtol, Raya Fidel, and Barbara H. Kwasnik, 39–56. Silver Spring, MD: American Society for Information Science.
- Bowker, Lynne. 1997a. "Multidimensional Classification of Concepts and Terms." In *Handbook of Terminology Management Vol. 1*, edited by Sue Ellen Wright, and Gerhard Budin, 133–143. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.htm1.16bow>
- Bowker, Lynne. 1997b. "You Say 'Flatbed Colour Scanner', I Say 'Colour Flatbed Scanner': A Descriptive Study of the Influence of Multidimensionality on Term Formation and Use with Special Reference to the Subject Field of Optical Scanning Technology." *Terminology* 4 (2): 275–302. <https://doi.org/10.1075/term.4.2.o4bow>
- Bowker, Lynne. 1998. "Variant Terminology: Frivolity or Necessity." In *8th EURALEX International Congress Proceedings*, Liège, Belgium, edited by Thierry Fontenelle, Philippe Hilgsmann, Archibald Michiels, André Martin, and Siegfried Theissen, 487–495. Liège: Université de Liège.
- Bowker, Lynne. 2011. "Off the Record and on the Fly." In *Corpus-based Translation Studies: Research and Applications*, edited by Alet Kruger, Kim Wallmach, and Jeremy Munday, 211–236. London/New York: Continuum.
- Bowker, Lynne, and Shane Hawkins. 2006. "Variation in the Organization of Medical Terms: Exploring Some Motivations for Term Choice." *Terminology* 12 (1): 79–110. <https://doi.org/10.1075/term.12.1.o5bow>
- Bowker, Lynne, and Timothy C. Lethbridge. 1994. "CODE4: Applications for Managing Classification Schemes." In *Advances in Classification Research: Proceedings of the 5th ASIS SIG/CR Classification Research Workshop*, edited by Raya Fidel, Clare Beghtol, Barbara H. Kwasnik, and Philip J. Smith, 17–32. Silver Spring, MD: American Society for Information Science (ASIS).
- Bowker, Lynne, and Ingrid Meyer. 1993. "Beyond 'Textbook' Concept Systems: Handling Multidimensionality in a New Generation of Term Bank." In *Proceedings of the Third International Congress on Terminology and Knowledge Engineering*, edited by Klaus-Dirk Schmitz, 123–134. Frankfurt: Indeks Verlag.
- Bowker, Lynne, and Jennifer Pearson. 2002. *Working with Specialized Language: A Practical Guide to Using Corpora*. London/New York: Routledge. <https://doi.org/10.4324/9780203469255>
- Boyd, Richard. 1993. "Metaphor and Theory Change: What is 'Metaphor' a Metaphor for?" In *Metaphor and Thought*, edited by Andrew Ortony, 481–533. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139173865.023>
- Bradford, Cristine, and Thomas Dana. 1996. "Exploring Science Teacher Metaphorical Thinking: A Case Study of a High School Science Teacher." *Journal of Science Teacher Education* 7: 197–211. <https://doi.org/10.1007/BF00117035>
- Brdar, Mario, Rita Brdar-Szabó, and Benedikt Perak. 2019. "Metaphor Repositories and Cross-Linguistic Comparison: Ontological Eggs and Chickens." In *Metaphor and Metonymy in the Digital Age*, edited by Marianna Bolognesi, Mario Brdar, and Kristina Despot, 225–252. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/milcc.8.10brd>

- Brodt, Garvin, Markus Werkle-Bergner, and Yee Lee Shing. 2013. "The Influence of Prior Knowledge on Memory: A Developmental Cognitive Neuroscience Perspective." *Frontiers in Behavioral Neuroscience* 7: 139. <https://doi.org/10.3389/fnbeh.2013.00139>
- Brown, Theodore. 2003. *Making Truth. Metaphor in Science*. Illinois: University of Illinois Press.
- Brunet, Pauline, Olivier Ferret, and Ludovic Tanguy. 2020. "Which Dependency Parser to Use for Distributional Semantics in a Specialized Domain?" In *Proceedings of the 6th International Workshop on Computational Terminology (COMPUTERM 2020)*, edited by Béatrice Daille, Kyo Kageura, and Ayla Rigouts Terryn, 26–36. Luxembourg: European Language Resources Association.
- Bruni, Elia, Nam Khanh Tran, and Marco Baroni. 2014. "Multimodal Distributional Semantics." *Journal of Artificial Intelligence Research* 49 (1): 1–47. <https://doi.org/10.1613/jair.4135>
- Budin, Gerhard. 1998. "The Wüster Archive – A Special Node in a European Digital Archive Network." In *Infoterm, Proceedings of the International Conference on Professional Communication and Knowledge Transfer*, edited by Erhard Oeser, and Christian Galinski, 169–173. Vienna: TermNet.
- Budin, Gerhard. 2001. "A Critical Evaluation of the State of the Art of Terminology Theory." *Terminology Science and Research* 12 (1–2): 7–23.
- Budin, Gerhard. 2006. "Terminology Studies in German-speaking Communities." In *Modern Approaches to Terminological Theories and Applications*, edited by Heribert Picht, 91–106. Bern: Peter Lang.
- Budin, Gerhard. 2007. "L'apport de la philosophie autrichienne au développement de la théorie de la terminologie : ontologie, théories de la connaissance et de l'objet." *Langages* 168: 11–23. <https://doi.org/10.3917/lang.168.0011>
- Budin, Gerhard. 2018. "Wüster, Schmitz und die Folgen, Entwicklungslinien der Terminologielehre und ihre Anwendungsbereiche." In *Verschmitzt! Von Terminologie und Terminologen. Festschrift für Klaus-Dirk Schmitz*, edited by Barbara Ahrens, Lisa Link, Ute B. Schilly, and Ursula Wienen, 89–107. Berlin: Frank & Timme.
- Budin, Gerhard, and Erhard Oeser. 1995. "Controlled Conceptual Dynamics: From 'Ordinary Language' to Scientific Terminology – and Back." *Terminology Science and Research* 6 (2): 3–17.
- Buendia-Castro, Miriam, Silvia Montero-Martínez, and Pamela Faber. 2014. "Verb Collocations and Phraseology in EcoLexicon." *Yearbook of Phraseology* 5 (1): 57–94. De Gruyter.
- Bühler, Hildegund. 1982a. *Aus dem wissenschaftlichen Nachlass von Eugen Wüster (1898–1977), Unvollendete und unveröffentlichte Arbeiten auf dem Gebiet der Terminologie*, Unter Mitarbeit von Ulrike Seipelt, 4 vols. Vienna, typescript.
- Bühler, Hildegund. 1982b. "General Theory of Terminology and Translation Studies." *Meta* 27 (4): 425–431. <https://doi.org/10.7202/002174ar>
- Bühler, Hildegund. 2005. "Translation(wissenschaft) und Terminologie(wissenschaft) im akademischen Fächerkanon." In *Fluctuat nec mergitur Translation und Gesellschaft, Festschrift Annemarie Schmid*, edited by Peter Sandrini, 71–84. Frankfurt: Peter Lang Verlag.
- Buitelaar, Paul, and Philip Cimiano. 2008. *Ontology Learning and Population: Bridging the Gap between Text and Knowledge*. Amsterdam: IOS Press.
- Buitelaar, Paul, Philipp Cimiano, Peter Haase, and Michael Sintek, M. 2009. "Towards linguistically grounded ontologies." In *Proceedings of the 6th European Semantic Web Conference (ESWC09)*, 111–125. Heraklion, Greece. https://doi.org/10.1007/978-3-642-02121-3_12
- Buitelaar, Paul, Thierry Declerck, Anette Frank, Stefania Racioppa, Malte Kiesel, Michael Sintek, Ralf Engel, Massimo Romanelli, Daniel Sonntag, Berenike Loos, Vanessa Micelli, Robert Porzel, and Philipp Cimiano. 2006. "LingInfo: Design and Applications of a Model for the Integration of Linguistic Information in Ontologies." In *Proceedings of the OntoLex 2006*

- Workshop: Interfacing Ontologies and Lexical Resources for Semantic Web Technologies*. Genova, Italy. (https://pub.uni-bielefeld.de/download/2497539/2525178/Buitelaar_Ling-Info_Desig_1.pdf). Accessed 7 December 2021.
- Buono, Maria Pia, Philipp Cimiano, Mohammad Fazleh Elahi, and Frank Grimm. 2020. "Termè-à-LLOD: Simplifying the Conversion and Hosting of Terminological Resources as Linked Data." In *Proceedings of the 7th Workshop on Linked Data in Linguistics*. Marseille, France. (<https://aclanthology.org/2020.ldl-1.5.pdf>). Accessed 7 December 2021.
- Burenhult, Niclas, and Stephen C. Levinson. 2008. "Language and Landscape: A Crosslinguistic Perspective." *Language Sciences* 30: 135–150. <https://doi.org/10.1016/j.langsci.2006.12.028>
- Bürger, Thomas, and Elena Simperl. 2008. "Measuring the benefits of ontologies." In *Proceedings of On the Move to Meaningful Internet Systems Workshop (OTM 2008)*, edited by Robert Meersman, Zahir Tari, and Pilar Herrero, 584–594. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-540-88875-8_82
- Busse, Dietrich. 2012. *Frame-Semantik. Ein Kompendium*. Berlin/New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110269451>
- Buttigieg, Pier Luigi, Norman Morrison, Barry Smith, Christopher J. Mungall, and Suzanna E. Lewis. 2013. "The Environment Ontology: Contextualising Biological and Biomedical Entities." *Journal of Biomedical Semantics* 4 (43): 4–43. <https://doi.org/10.1186/2041-1480-4-43>
- Buttigieg, Pier Luigi, Evangelos Pafilis, Suzanna E. Lewis, Mark P. Schildhauer, Ramona L. Wallis, and Christopher J. Mungall. 2016. "The Environment Ontology in 2016: Bridging Domains with Increased Scope, Semantic Density, and Interoperation." *Journal of Biomedical Semantics* 7, article 57. <https://doi.org/10.1186/s13326-016-0097-6>
- ## C
- Caballero, Rosario. 2006. *Re-Viewing Space: Figurative Language in Architects' Assessment of Built Space*. Berlin/New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110893892>
- Caballero, Rosario. 2012. "The Role of Metaphor in Tennis Reports and Forums." *Text and Talk* 32 (6): 703–726. <https://doi.org/10.1515/text-2012-0033>
- Cabezas-García, Melania. 2019. "Los compuestos nominales en Terminología: formación, traducción y representación." PhD diss., University of Granada.
- Cabezas-García, Melania, and Pamela Faber. 2017a. "A Semantic Approach to the Inclusion of Complex Nominals in English Terminographic Resources." In *Computational and Corpus-Based Phraseology*, edited by Ruslan Mitkov, 145–159. Berlin: Springer. https://doi.org/10.1007/978-3-319-69805-2_11
- Cabezas-García, Melania, and Pamela Faber. 2017b. "Exploring the Semantics of Multi-Word Terms by Means of Paraphrases." In *Temas actuales de Terminología y estudios sobre el léxico*, edited by Miguel Ángel Candel-Mora, and Chelo Vargas-Sierra, 193–217. Granada: Comares.
- Cabezas-García, Melania, and Pamela Faber. 2017c. "The Role of Micro-contexts in Noun Compound Formation." *Neologica* 11: 101–118.
- Cabezas-García, Melania, and Pamela Faber. 2018. "Phraseology in Specialized Resources: An Approach to Complex Nominals." *Lexicography* 5 (1): 55–83. <https://doi.org/10.1007/s40607-018-0046-x>
- Cabezas-García, Melania, and Juan Carlos Gil-Berrozpe. 2018. "Semantic-based Retrieval of Complex Nominals in Terminographic Resources." In *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*, edited by Jaka Čibej, Vojko Gorjanc, Iztok Kosem, and Simon Krek, 269–281. Ljubljana: Ljubljana University Press.

- Cabezas-García, Melania, and Pilar León-Araúz. 2018. "Towards the Inference of Semantic Relations in Complex Nominals: A Pilot Study." In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, 2511–2518. Myazaki, Japan.
- Cabezas-García, Melania, and Pilar León-Araúz. 2021. "Procedimiento para la traducción de términos compuestos con la ayuda de corpus." In *Sistemas fraseológicos en contraste: enfoques computacionales y de corpus*. Granada: Comares.
- Cabré, María Teresa. 1995. "On Diversity and Terminology." *Terminology* 2 (1): 1–16. <https://doi.org/10.1075/term.2.1.02cab>
- Cabré, María Teresa. 1996. "Presentació." In *Terminologia: Selecció de textos d'E. Wüster*, edited by Maria Teresa Cabré, 5–7. Barcelona: Servei de Llengua catalana.
- Cabré, María Teresa. 1998a. "El discurs especialitzat o la variació funcional determinada per la temàtica: noves perspectives." *Caplletra. Revista internacional de filologia* 25: 173–194.
- Cabré, María Teresa. 1998b. *La terminologie : théorie, méthode et applications*. Translated by Monique C. Cormier and John Humbley. Ottawa/Paris: Ottawa University Press/Armand Colin.
- Cabré, María Teresa. 1999a. "Elementos para una teoría de base comunicativa y otros artículos." In *La terminología: representación y comunicación*, edited by Maria Teresa Cabré, 69–92. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra. <https://doi.org/10.1075/tlrp.1>
- Cabré, María Teresa. 1999b. *La terminología: Representación y comunicación*. Barcelona: IULA. Universitat Pompeu Fabra. <https://doi.org/10.1075/tlrp.1>
- Cabré, María Teresa. 1999c. *Terminology, Theory, Method, and Applications*, edited by Juan C. Sager. Translated by Janet Ann DeCesaris. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.1>
- Cabré, María Teresa. 2000. "Terminologie et linguistique: la théorie des portes." *Terminologies nouvelles* 21, *Terminologie et diversité culturelle*: 10–15.
- Cabré, María Teresa. 2003. "Theories of Terminology: Their Description, Prescription, and Explanation." *Terminology* 9 (2): 163–199.
- Cabré, María Teresa. 2005. "A Terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro." *Revista Debate Terminológico* 1. (<http://seer.ufrgs.br/riterm/article/view/21286>). Accessed 7 December 2021.
- Cabré, María Teresa. 2006. "From Terminological Data Banks to Knowledge Databases: The Text as the Starting Point." In *Lexicography, Terminology, and Translation. Text-based Studies in Honour of Ingrid Meyer*, edited by Lynne Bowker, 93–106. Ottawa: University of Ottawa Press.
- Cabré, María Teresa. 2008. "El principio de poliedricidad: la articulación de lo discursivo, lo cognitivo y lo lingüístico en Terminología (I)." *IBÉRICA* 16: 9–36.
- Cabré, María Teresa. Forthcoming. *Terminology: Cognition, Language and Communication*, Amsterdam/Philadelphia: John Benjamins.
- Cabré, María Teresa, Carme Bach, Rosa Estopà, Judit Feliu, Gerardo Martínez, and Jorge Vivaldi. 2004. "The GENOMA-KB project: towards the integration of concepts, terms, textual corpora and entities." In *Proceedings of the 4th International Conference on Language Resources and Evaluation 1*: 87–90. Lisbon, Portugal.
- Cabré, María Teresa, Rosa Estopà, and Chelo Vargas-Sierra (eds). 2012. *Neology in Specialized Communications*. Special Issue of *Terminology* 18 (1): 1–8. <https://doi.org/10.1075/term.18.1.01int>
- Cabré, María Teresa, and Carles Tebé. 2005. "El trabajo terminológico multilingüe de enriquecimiento lingüístico: una propuesta metodológica." *Revista española de lingüística aplicada* 1: 19–41.

- Campo, Ángela. 2012. "The Reception of Eugen Wüster's Work and the Development of Terminology." PhD diss., Université de Montréal.
- Candel, Danielle. 2004. "Wüster par lui-même." In *Des fondements théoriques de la terminologie*, Cahiers du C.I.E.L. 2004–2005 :15–32.
- Candel, Danielle. 2007. "Terminologie de la terminologie. Métalangage et reformulation dans l'Introduction à la terminologie générale et à la lexicographie terminologique d'E. Wüster." *Langages* 168: 66–81. <https://doi.org/10.3917/lang.168.0066>
- Candel, Danielle. 2017. "Eugen Wüster (1898–1977). 'Terminologie et variation'," In *Terminologie et ontologie : théorie et applications. Actes de la conférence TOTh 2012*, 21–37. Chambéry: Université Savoie Mont-Blanc.
- Candel, Danielle, and François Gaudin (eds). 2006. *Aspects diachroniques du vocabulaire*. Rouen: Publications des Universités de Rouen et du Havre.
- Caple, Helen, Monika Bednarek, and Lawrence Anthony. 2018. "Using Kaleidographic to Visualize Multimodal Relations within and across Texts." *Visual Communication* 17 (4): 461–474. <https://doi.org/10.1177/1470357218789287>
- Carné, Louis de. 1835. "Une question de politique et d'histoire." *Le correspondant* 36: 9.
- Carrió, María Luisa, and Miguel Ángel Candel. 2012. "Corpus Analysis: a Pragmatic Perspective on Term Variation." *Revista Española de Lingüística Aplicada* 25 (1): 33–50.
- Casals, Sergi. 2006. "Reflexiones sobre la variación terminológica del español científico ilustradas con el caso del término inglés *delusion*." *Panace@: Revista de Medicina, Lenguaje y Traducción* 7 (24): 222–227.
- Cassani, Giovanni, and Alessandro Lopopolo. 2015. "Multimodal Distributional Semantics Models and Conceptual Representations in Sensory Deprived Subjects." Poster presentation at *Computational Linguistics in the Netherlands (CLIN 2015)*, Antwerp, Belgium. <https://doi.org/10.13140/RG.2.1.5096.9369>
- Cerbah, Farid. 2000. "Exogeneous and Endogeneous Approaches to Semantic Categorisation of Unknown Technical Terms." In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*, 145–151. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/990820.990842>
- Ceusters, Werner, and Barry Smith. 2006. "A Realism-based Approach to the Evolution of Bio-medical Ontologies," In *Proceedings of the Annual AMIA Symposium*, 121–125. Washington, DC: American Medical Informatics Association.
- Charaudeau, Patrick, and Dominique Maingueneau (eds). 2002. *Dictionnaire d'analyse du discours*. Paris: Le Seuil.
- Charette, Peter, Neal H. Hooker, and John L. Stanton. 2015. "Framing and Naming: A Process to Define a Novel Food Category." *Food Quality and Preference* 40: 147–151. <https://doi.org/10.1016/j.foodqual.2014.09.015>
- Charteris-Black, Jonathan. 2014. *Analysing Political Speeches: Rhetoric, Discourse and Metaphor*. Basingstoke/ New York: Palgrave-MacMillan. <https://doi.org/10.1007/978-1-137-36833-1>
- Chen, Long-Huei, and Kyo Kageura. 2020. "Multilingualization of Medical Terminology: Semantic and Structural Embedding Approaches." In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, 4157–4166. Paris: European Language Resources Association.
- Chiocchetti, Elena, and Natascia Ralli (eds). 2013. *Guidelines for Collaborative Legal/Administrative Terminology Work*. Bolzano/Bozen: LISE.
- Chukwu, Uzoma E. 1996. "Science, dénomination et partage du pouvoir : le cas des éponymes." *Meta* 41 (4): 590–603. <https://doi.org/10.7202/002242ar>

- Church, Ken W. 2011. "A Pendulum Swung too Far." *Linguistic Issues in Language Technology* 6: 1–27. <https://doi.org/10.33011/ilt.v6i.1245>
- Ciapuscio, Guiomar Elena. 2003. *Textos especializados y terminología*. Barcelona: IULA-UPF.
- Ciapuscio, Guiomar Elena. 2007. "Genres et familles de genres : apports pour l'acquisition de la compétence générique dans le domaine académique." *Éla. Études de linguistique appliquée* 148: 405–416. <https://doi.org/10.3917/ela.148.0405>
- Cimiano, Philipp. 2006. *Ontology Learning and Population from Text. Algorithms, Evaluation and Applications*. Berlin/Heidelberg: Springer.
- Cimiano, Philipp, Christian Chiarcos, John McCrae, and Jorge Gracia (eds). 2020. *Linguistic Linked Data: Representation, Generation and Applications*. Berlin/Heidelberg: Springer. <https://doi.org/10.1007/978-3-030-30225-2>
- Cimiano, Philipp, Cristian Chiarcos, John P. McCrae, and Jorge Garcia. 2020. "Linguistic Linked Data Cloud." In *Linguistic Linked Data: Representation, Generation and Applications*, 29–41. New York: Springer. https://doi.org/10.1007/978-3-030-30225-2_3
- Cimiano, Philipp, Peter Haase, Matthias Herold, Matthias Mantel, and Paul Buitelaar. 2007. "LexOnto: A Model for Ontology Lexicons for Ontology-based NLP." In *Proceedings of the OntoLex07 Workshop at the ISWC07*. Busan, Korea (<https://pub.uni-bielefeld.de/download/2497432/2524813>). 7 December 2021.
- Cimiano, Philipp, Aleksander Pivk, Lars Schmidt-Thieme, and Steffen Staab. 2005. "Learning Taxonomic Relations from Heterogeneous Evidence." In *Ontology learning from text: Methods, applications and evaluation*, 59–73. Amsterdam: IOS Press.
- Cimiano, Philipp, and Johanna Wenderoth. 2005. "Learning Qualia Structures from the Web." In *ACL workshop on Deep Lexical Acquisition*, 28–37. Michigan, USA. <https://doi.org/10.3115/1631850.1631854>
- Cimiano, Philipp, and Johanna Wenderoth. 2007. "Automatic Acquisition of Ranked Qualia Structures from the Web." In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL07)*, 888–895. Prague, Czech Republic.
- Cimiano, Philipp, John McCrae, Paul Buitelaar, and Elena Montiel-Ponsoda. 2012. "On the Role of Senses in the Ontology-Lexicon." In *New Trends of Research in Ontologies and Lexical Resources*, edited by Alessandro Oltramari, Piek Vossen, Lu Qin, and Eduard Hovy, 43–62. Berlin/Heidelberg: Springer.
- Cimiano, Philipp, John McCrae, Víctor Rodríguez-Doncel, Tatiana Gornostay, Asunción Gomez-Pérez, Benjamin Siemoneit, and Andis Lagzdins. 2015. "Linked Terminologies: Applying Linked Data Principles to Terminological Resources." In *Proceedings of the fourth biennial conference on electronic lexicography (eLex 2015)*, 504–517. Sussex, United Kingdom.
- Clark, Stephen. 2015. "Vector Space Models of Lexical Meaning." In *Handbook of Contemporary Semantics, 2nd edition*, edited by Shalom Lappin, and Chris Fox, 493–522. Oxford: Wiley-Blackwell. <https://doi.org/10.1002/9781118882139.ch16>
- Clas, André. 2001. "Review of Temmerman 2000." *Meta* 46 (3): 582–585. <https://doi.org/10.7202/002265ar>
- Claveau, Vincent, and Ewa Kijak. 2015. "Thésaurus distributionnels pour la recherche d'information et vice-versa." *Revue des Sciences et Technologies de l'Information – Série Document Numérique* 18 (2): 101–121. <https://doi.org/10.3166/dn.18.2-3.101-121>
- Cohen, S. Marc and C. D. C. Reeve. 2020. "Aristotle's Metaphysics." In *The Stanford Encyclopedia of Philosophy (Winter 2020 Edition)*, edited by Edward N. Zalta. Stanford: Stanford University (<https://plato.stanford.edu/archives/win2020/entries/aristotle-metaphysics/>). Accessed 7 December 2021.

- Cohen, Trevor, and Dominic Widdows. 2009. "Empirical Distributional Semantics: Methods and Biomedical Applications." *Journal of Biomedical Informatics* 42 (2): 390–405.
<https://doi.org/10.1016/j.jbi.2009.02.002>
- Cole, Waine. 1991. "Descriptive Terminology: Some Theoretical Implications." *Meta* 36 (1): 16–22.
<https://doi.org/10.7202/002347ar>
- Collet, Tanja. 2004. "What's a Term? An Attempt to Define the Term within the Theoretical Framework of Text Linguistics." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 3: 99–111.
- Collonval, Sophie. 2021. *Les nouveaux arrivants peuvent-ils entrer ? De l'expérience de la vulnérabilité à la quête de reconnaissance d'une place en tant que soignants dans une équipe hospitalière. Regard sociolinguistique sur une transition professionnelle*. Namur: Université de Namur.
- Condamines, Anne. 1993. "Un exemple d'utilisation de connaissances de sémantique lexicale: acquisition semiautomatique d'un vocabulaire de spécialité." *Cahiers de lexicologie* 62 (1): 25–65.
- Condamines, Anne. 1995. "Terminology: New Needs, New Perspectives." *Terminology* 2(2): 219–238. <https://doi.org/10.1075/term.2.2.03con>
- Condamines, Anne. 2000. "Les bases théoriques du groupe toulousain 'Sémantique et corpus': ancrages et perspectives." *Cahiers de grammaire* 25: 5–28.
- Condamines, Anne. 2002. "Corpus Analysis and Conceptual Relation Patterns." *Terminology* 8 (1): 141–162. <https://doi.org/10.1075/term.8.1.07con>
- Condamines, Anne. 2003. "Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques." *Habilitation à Diriger des Recherches*, l'Université Toulouse Le Mirail.
- Condamines, Anne. 2005. "Linguistique de corpus et terminologie." *Langages* 39 (157): 36–47.
<https://doi.org/10.3406/lgge.2005.973>
- Condamines, Anne. 2008a. "Peut-on prévenir le risque langagier dans la communication écrite?" *Langage et société* 3 (125): 77–97. <https://doi.org/10.3917/l.s.125.0077>
- Condamines, Anne. 2008b. "Taking Genre into Account when Analysing Conceptual Relation Patterns." *Corpora* 3 (2): 115–140. <https://doi.org/10.3366/E1749503208000129>
- Condamines, Anne. 2010. "Variations in Terminology. Application to the Management of Risks Related to Language Use in the Workplace." *Terminology* 16 (1): 30–50.
<https://doi.org/10.1075/term.16.1.02con>
- Condamines, Anne. 2014. "How can Linguistics Help to Structure a Multidisciplinary Neo-Domain such as Exobiology?" *BIO Web of Conferences* 2, 06001.
<https://doi.org/10.1051/bioconf/20140206001>
- Condamines, Anne. 2017. "The Emotional Dmension in Terminological Variation: The Example of Transitivity of the Locative Complement in Fishing." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, John Humbley, Aline Francoeur, and Aurélie Picton, 11–30. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/tlrp.18.01con>
- Condamines, Anne. 2018a. "La linguistique appliquée en entreprise : une linguistique ergonomique à l'intersection de la linguistique de corpus, de la psycholinguistique et de l'ergonomie." *Études de linguistique appliquée* 190 (2): 205–216. <https://doi.org/10.3917/ela.190.0205>
- Condamines, Anne. 2018b. "Nouvelles perspectives pour la terminologie textuelle." In *Terminology and Discourse*, edited by Jana Altmanova, Maria Centrella, and Katherine E. Russo, 93–112. Bern/New York: Peter Lang.
- Condamines, Anne. 2018c. "Terminological Knowledge Bases." In *The Routledge Handbook of Lexicography*, edited by Pedro Fuertes, 335–349. London: Routledge.

- Condamines, Anne. 2021. "Towards an Ergonomic Linguistics. Application to the Design of Controlled Natural Languages." *International Journal of Applied Linguistics* 31: 18–30. <https://doi.org/10.1111/ijal.12313>
- Condamines, Anne, Julie Humbert-Droz, and Aurélie Picton. 2021. "Néologie par détermination : méthode de repérage et catégorisation en corpus dans le domaine de la physique des particules." In *La néologie des langues romanes: nouvelles dynamiques et enjeux*, edited by Maria Bélen Villar Díaz, José Carlos de Hoyos, Pascaline Dury, Julie Makri-Morel, and Vincent Renner, 287–304. Bern: Peter Lang.
- Condamines, Anne, Amélie Josselin-Leray, Cécile Fabre, Luce Lefeuvre, Aurélie Picton, and Josette Rebeyrolle. 2013. "Using Comparable Corpora to Characterize Knowledge-Rich Contexts for Various Kinds of Users: Preliminary Steps." *Procedia – Social and Behavioral Sciences* 95: 581–586. <https://doi.org/10.1016/j.sbspro.2013.10.685>
- Condamines, Anne, and Jean-Paul Narcy-Combes. 2015. "La linguistique appliquée comme science située." In *Cultures de Recherche En Linguistique Appliquée*, edited by Francis Carton, Jean-Paul Narcy-Combes, Marie-Françoise Narcy-Combes, and Danelle Toffoli, 209–229. Paris: Riveneuve.
- Condamines, Anne, and Aurélie Picton. 2012. "Des communiqués de presse du Cnes à la presse généraliste. Vers un observatoire de la diffusion des termes." In *Langues de spécialité et néologie*, edited by Pascaline Dury, José Carlos de Hoyos, Julie Makri-Morel, François Maniez, Vincent Renner, and Maria Belén Villar Díaz, 141–161. Lyon: CRTT.
- Condamines, Anne, and Josette Rebeyrolle. 1997. "Point de vue en langue spécialisée." *Meta* 42 (1): 174–84. <https://doi.org/10.7202/002359ar>
- Condamines, Anne, and Josette Rebeyrolle. 2001. "Searching for and Identifying Conceptual Relationships via a Corpus-based Approach to a Terminological Knowledge Base (CTKB): Method and Results." In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, 127–148. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/nlp.2.07con>
- Condamines, Anne, Josette Rebeyrolle, and Anne Soubeille. 2004. "Variation de la terminologie dans le temps: une méthode linguistique pour mesurer l'évolution de la connaissance en corpus." In *Proceedings of the EURALEX International Congress*, Lorient, France, July 6–10, 2004, edited by Geoffrey Williams, and Sandra Vessier, 547–557. Lorient: Université de Bretagne-Sud.
- Copara, Jenny, Julien Knafo, Nona Naderi, Claudia Moro, Patrick Ruch, and Douglas Teodoro. 2020. "Contextualized French Language Models for Biomedical Named Entity Recognition." In *Actes de la 27^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2020) Atelier Défi Fouille de Textes*, Nancy, France, June 2020, edited by Rémi Cardon, Natalia Grabar, Cyril Grouin, and Thierry Hamon, 36–48. Nancy: ATALA /AFCP.
- Corbeil, Jean Claude. 1980. *L'aménagement linguistique du Québec*. Collection Langue et Société 3. Montreal: Guérin.
- Corbeil, Jean-Claude. 2003. "La contribution de Josette Rey-Debove et d'Alain Rey à l'aménagement de la langue au Québec." In *Les dictionnaires Le Robert: genèse et évolution*, edited by Monique Cormier, Aline Francoeur, and Jean-Claude Boulanger, 263–279. Montreal: Presses universitaires de Montréal. <https://doi.org/10.4000/books.pum.13871>
- Cornelissen, Joep, and Mario Kafourous. 2008. "The Emergent Organization: Primary and Complex Metaphors in Theorizing about Organizations." *Organization Studies* 29 (7): 957–978. <https://doi.org/10.1177/0170840608090533>

- Corpas Pastor, Gloria, and Miriam Seghiri Domínguez. 2010. "Size Matters: A Quantitative Approach to Corpus Representativeness." In *Lengua, traducción, recepción: en honor de Julio César Santoyo*, edited by Rosa Rabadán, Marisa Fernández López, and Trinidad Guzmán González, 111–145. Vol. 1. León: Publicaciones Universidad de León.
- Cortés de los Ríos, Enriqueta. 2001. "The Teaching of Metaphor in the Advertising Discourse of Business Newspapers." *Ibérica* 3: 91–106.
- Cossette, Mychèle. 1976. "Description de l'acte de terminologie." In *Handbook on Terminology*, Ottawa, Canada: Department of the Secretary of State of Canada (Terminology and Linguistic Services Directorate).
- Coutherut, Margaux. 2017. "L'électronicien francophone devant le manuel d'utilisation d'appareil rédigé ou à rédiger en anglais." *ASP la revue du GERAS* 71: 69–92.
<https://doi.org/10.4000/asp.4971>
- Croft, William, and Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511803864>
- Cruse, Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Cruse, Alan. 1995. "Polysemy and Related Phenomena from a Cognitive Linguistics Viewpoint." In *Computational Lexical Semantics*, edited by Patrick Saint-Dizier, and Evelyn Viegas, 33–49. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511527227.004>

D

- Daconta, Michel C., Leo J. Obrst, and Kevin Smith. 2005. *The Semantic Web: A Guide to the Future of XML, Web Services and Knowledge Management*. Indianapolis: Wiley.
- Dahlberg, Ingetraut. 1978. "A Referent-Oriented, Analytical Concept Theory for INTERCONCEPT." *International Classification* 5(3): 142–151.
<https://doi.org/10.5771/0943-7444-1978-3-142>
- Dahlberg, Ingetraut. 1981. "Conceptual Definitions for INTERCONCEPT." *International Classification* 8(1): 16–22. <https://doi.org/10.5771/0943-7444-1981-1-16>
- Daille, Béatrice. 2001. "Qualitative Terminology Extraction." In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, 149–166. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/nlp.2.08dai>
- Daille, Béatrice. 2017. *Term Variation in Specialised Corpora: Characterisation, Automatic Discovery and Applications*. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/tlrp.19>
- Daille, Beatrice, Samuel Dufour-Kowalski, and Emmanuel Morin. 2004. "French-English Multi-Word Term Alignment Based on Lexical Context Analysis." In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, edited by María Teresa Lino, Maria Francisca Xavier, Fatima Ferreira, Rute Costa, and Raquel Silva, 919–922. Lisbon: European Language Resources Association.
- Daille, Béatrice, Benoît Habert, Christian Jacquemin, and Jean Royauté. 1996. "Empirical Observation of Term Variations and Principles for their Description." *Terminology* 3 (2): 197–257.
<https://doi.org/10.1075/term.3.2.02dai>
- Dalla-Zuanna, Jean-Marc, and Christopher Kurz. 2019. "Fehler, Fehlerkategorien, und Fehlerbewertungen im Zeitalter der Industrie 4.0." In *Übersetzen und Dolmetschen 4.0: Neue Wege im digitalen Zeitalter* 102 3. *Internationale Fachkonferenz des BDÜ*, Bonn, Germany.

- De Bessé, Bruno. 1974. "Observation de la synonymie dans la terminologie métallurgie." In *La normalisation linguistique. Actes du colloque international de terminologie*, Lac Delage, Québec, October 16–19, 1973, edited by Monique Héroux, 35–47. Quebec: Éditeur officiel du Québec.
- De Bessé, Bruno, Blaise Nkwenti-Azeh, and Juan C. Sager. 1997. "Glossary of Terms Used in Terminology." *Terminology* 4 (1): 117–156. <https://doi.org/10.1075/term.4.1.o8bes>
- Dechamps, Cristina. 2013. "L'enseignement/apprentissage du français juridique: une proposition de démarche terminologique et didactique." *Equivalences* 40 (1–2): 191–124. <https://doi.org/10.3406/equiv.2013.1387>
- Deignan, Alice. 2007. "'Image' Metaphors and Connotations in Everyday Language." *Annual Review of Cognitive Linguistics* 5: 173–192. <https://doi.org/10.1075/arcl.5.o8dei>
- Del Tredici, Marco, Raquel Fernández, and Gemma Boleda. 2019. "Short-Term Meaning Shift: A Distributional Exploration." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, edited by Jill Burstein, Christy Doran, and Thamar Solorio, 2069–2075. Stroudsburg: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1210>
- Delaveau, Pierre. 2007. *La mémoire des mots en médecine, pharmacie et sciences*. Paris: Louis Pariente.
- Delavigne, Valérie. 1995. "Approche socioterminologique des discours du nucléaire." *Meta* 44 (2): 308–319. <https://doi.org/10.7202/002506ar>
- Delavigne, Valérie. 1996. "Problèmes d'enquêtes en socioterminologie." In *Actes des Secondes journées internationales de Terminologie*, 163–178. Le Havre, France.
- Delavigne, Valérie. 2001. "Les mots du nucléaire. Contribution socioterminologique à une analyse des discours de vulgarisation." PhD diss., Université de Rouen.
- Delavigne, Valérie. 2006. "La formation du vocabulaire de la physique nucléaire: quelques jalons." In *Aspects diachroniques du vocabulaire*, edited by Danielle Candel, and François Gaudin, 89–109. Rouen: Publications des Universités de Rouen et du Havre.
- Delavigne, Valérie. 2008. "Construire un dictionnaire d'oncologie pour les patients: aspects méthodologiques." In *Corpus et dictionnaires de langues de spécialité*, edited by François Maniez, Pascaline Dury, Nathalie Arlin, and Claire Rougemont, 153–173. Lyon: Presses Universitaires de Lyon.
- Delavigne, Valérie. 2012a. "Peut-on traduire les mots des experts? Un dictionnaire pour les patients atteints de cancer." In *Dictionnaires et traduction*, edited by Michaela Heinz, 233–266. Berlin: Frank & Timme.
- Delavigne, Valérie (ed.). 2012b. *Dictionnaire*. Institut national du cancer. (<http://www.e-cancer.fr/Dictionnaire>). Accessed 7 December 2021.
- Delavigne, Valérie. 2013. "Quand le patient devient expert: usages des termes dans les forums médicaux." In *Terminologie et intelligence artificielle, Actes des dixièmes rencontres TIA 2013*. Paris, France. (<https://hal.archives-ouvertes.fr/hal-00924159>). Accessed 26 February 2021.
- Delavigne, Valérie. 2015. "Quand le linguiste devient acteur: vulgarisation médicale et prescription linguistique." In *Cultures de recherche en linguistique appliquée*, edited by Francis Carton, Jean-Paul Narcy-Combes, Marie-Françoise Narcy-Combes, and Denize Toffoli, 77–95. Paris: Riveneuve Éditions.
- Delavigne, Valérie. 2017. "Term Usage and Socioterminological Variation. The Impact of Social and Local Issues on the Movement of Terms." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francœur, John Humbley, and Aurélie Picton, 31–56. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.o2del>
- Delavigne, Valérie. 2019a. "Littératies en santé et forums de patients: des formes d'ergonomie discursive." *Études de linguistique appliquée* 195: 363–381. <https://doi.org/10.3917/ela.195.0363>

- Delavigne, Valérie. 2019b. "Les mots du cancer : le partage des termes pour l'élaboration d'une culture périmédicale." In *Comunicació, llenguatge i salut. Estratègies lingüístiques per millorar la comunicació amb el pacient*, Série Activitats, edited by Rosa Estopà, 153–183. Barcelona: IULA.
- Delavigne, Valérie. 2020a. "De l'(in)constance du métalinguistique dans un corpus de vulgarisation médicale." *Corela*, HS-31. <https://doi.org/10.4000/corela.11031>
- Delavigne, Valérie. 2020b. "Une analyse socioterminologique de forums de patients atteints de cancer : culture périmédicale et expertise." In *Les savoirs d'expérience en santé. Défis épistémologiques et enjeux identitaires*, edited by Sophie Arborio, Arnaud Halloy, Fabienne Hejoaka, and Emmanuelle Simon, 159–178. Nancy: Presses universitaires de Nancy.
- Delavigne, Valérie, Julien Carretier, Sylvie Brusco, Line Leichtnam-Dugarin, Marie Déchelette, and Thierry Philip. 2007. "Comprendre les mots du cancer : LexOnco, dictionnaire d'oncologie pour les personnes malades et leurs proches. Aspects méthodologiques." *Bulletin Du Cancer* 94 (4): 405–410.
- Delavigne, Valérie, and François Gaudin (eds). 1994. *Implantation des termes officiels. Terminologies nouvelles*. ACCT/Communauté Française de Belgique.
- Delavigne, Valérie, and François Gaudin. 1996. "A propos d'implantation terminologique. Questionner l'usage ou le sentiment linguistique ?" *Le questionnement social. Cahiers de linguistique sociale* (28–29): 131–40.
- Delavigne, Valérie, and Louis Guespin. 1992. "Nucléaire: risque et sécurité. Une recherche en socioterminologie." *Terminologies nouvelles* 7: 19–25.
- Delavigne, Valérie, and Dardo de Vecchi. 2016. "Socioterminologie et pragmatérminologie: Rencontres et Complémentarités." In *Terminologie et ontologie : Théories et applications*, edited by Christophe Roche, 141–156. Chambéry: Éditions de l'Université Savoie Mont Blanc.
- Delavigne, Valérie, and Dardo de Vecchi (eds). 2021. *Termes en discours. Entreprises et organisations*. Paris: Presses de la Sorbonne Nouvelle.
- Delisle, Jean. 2008. *La terminologie au Canada: Histoire d'une profession*. Brossard (Quebec): Linguattech.
- Demaecker, Christine. 2017. "Wine-tasting Metaphors and their Translation." *Terminology* 23 (1): 113–131. <https://doi.org/10.1075/term.23.1.05dem>
- De Pascale, Stefano. 2019. "Token-based Vector Space Models as Semantic Control in Lexical Sociolectometry." PhD diss., KU Leuven.
- Depecker, Loïc (ed.). 1997. *La mesure des mots: cinq études d'implantation terminologique*. Mont-Saint-Aignan: Publications de l'Université de Rouen.
- Depecker, Loïc. 2013. "Pour une ethnoterminologie." In *Dans tous les sens du terme*, edited by Jean Quirion, Loïc Depecker, and Louis-Jean Rousseau, 13–30. Ottawa: Ottawa University Press.
- Depierre, Amélie. 2007. "Souvent HAEMA varie ...: Les dérivés du grec HAEMA en anglais: Étude de cas de variation." *Terminology* 13 (2): 155–176. <https://doi.org/10.1075/term.13.2.03dep>
- "Le dépouillement en terminologie." 1980. *Terminology and Linguistic Services Directorate*, Ottawa, Canada: Department of the Secretary of State of Canada.
- Desmet, Isabel. 2001. "Terminologie, culture et société. Éléments pour une théorie variationniste de la terminologie et des langues de spécialité." *Cahiers du Rifal* 26: 3–13.
- Desmet, Isabel. 2006. "Variabilité et variation en terminologie et langues spécialisées : discours, textes et contextes." In *Actes des Septièmes journées scientifiques du Réseau 'Lexicologie, Terminologie, Traduction' de l'Agence Universitaire de la Francophonie: Mots, termes et contextes*, edited by Daniel Blampain, Philippe Thoiron, and Marc Van Campenhoudt, 235–247. Paris: Édition des archives contemporaines.
- Desmet, Isabel. 2007. "Terminologie, culture et société. Éléments pour une théorie variationniste de la terminologie et des langues de spécialité." *Cahiers du Rifal* 26 (1): 3–13.

- Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." arXiv:1810.04805 [cs.CL].
- Devylder, Simon, and Jordan Zlatev. 2020. "Cutting and Breaking Metaphors of the Self and the Motivation and Sedimentation Model." In *Figurative Meaning Construction in Thought and Language*, edited by Annalisa Baicchi, 254–281. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/ftl.9.11dev>
- Diki-Kidiri, Marcel. 1992. "Développement du sango en tant que langue officielle." In *Sango, The National Official Language of the Central African Republic*, edited by Helma Pasch, 25–31. Cologne: Rüdiger Köppe Verlag.
- Diki-Kidiri, Marcel. 1996. "Metaphor as a cultural basis for conceptualization and source of terminological neologism." In *Questions de glottopolitique: France, Afrique, Monde Méditerranéen*, edited by Amèle Achouri, Fabienne Leconte, Maman Mallam Garba, and Nicolas Tsekos, 87–193. Rouen: Université de Rouen-URA CNRS 1164.
- Diki-Kidiri, Marcel. 1998. "Question de méthode en terminologie en langues africaines." *Revue française de Linguistique appliquée : Terminologie. Nouvelles directions* 3 (2): 15–28. <https://doi.org/10.3917/rfla.032.0015>
- Diki-Kidiri, Marcel. 1999a. "Diversité dans l'observation du réel." In *Terminología y modelos culturales*, edited by Amelia Irazazábal Nerpell, Luís Fernando Lara Ramos et al., 61–66. Institut Universitari de Lingüística Aplicada (IULA), Pompeu Fabra University. Barcelona: Documenta Universitaria.
- Diki-Kidiri, Marcel. 1999b. "Terminologie pour le développement." In *Terminología y modelos culturales*, edited by Amelia Irazazábal Nerpell, Luís Fernando Lara Ramos et al., 67–74. Institut Universitari de Lingüística Aplicada (IULA), Pompeu Fabra University. Barcelona: Documenta Universitaria.
- Diki-Kidiri, Marcel. 1999c. "Le signifié et le concept dans la dénomination." *Meta* 4 (44): 573–581. <https://doi.org/10.7202/002566ar>
- Diki-Kidiri, Marcel. 2000. "Une approche culturelle de la terminologie." *Terminologies nouvelles* 21: 27–31.
- Diki-Kidiri, Marcel (ed.). 2008. *Vocabulaires scientifiques en langues africaines: pour une approche culturelle de la terminologie*. Paris: Karthala.
- Diki-Kidiri, Marcel. 2014. "Le vocabulaire juridique en Sängö, une application de la terminologie culturelle." In *Dynamics of Terminology: An Interdisciplinary Perspective on Monolingual and Multilingual Culture-bound Communication*, edited by Rita Temmerman, and Marc Van Campenhoudt, 99–110. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.16.06dik>
- Diki-Kidiri, Marcel, Hubert Joly, and Colette Murcia. 1981. *Guide de la néologie*. Paris: Conseil international de la langue française.
- Diki-Kidiri, Marcel, André Mololi, Gédéon Koyamofou, and Julien Kouanga. 2005. *Lexique de suite bureautique français-sängö et sängö-français*. Bangui: Institut de Linguistique Appliquée, Université de Bangui.
- Dinu, Georgiana, Stefan Thater, and Sören Laue. 2012. "A Comparison of Models of Word Meaning in Context." In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, edited by Eric Fosler-Lussier, Ellen Riloff, and Srinivas Bangalore, 611–615. Stroudsburg, PA: Association for Computational Linguistics.

- Dirven, René, Roslyn Frank, Tom Ziemke, and Jordan Zlatev. 2008. *Body, Language, and Mind*, vol. 2, *Social Situatedness*. Berlin/New York: Mouton de Gruyter.
- Dirven, René, and Francisco J. Ruiz de Mendoza. 2010. "Looking Back at 30 Years of Cognitive Linguistics." In *Cognitive Linguistics in Action. From Theory to Application and back*, edited by Elzbieta Tabakowska, Michal Choński, and Łukasz Wiraszka, 13–70. Berlin/New York: Mouton de Gruyter.
- Dolbec, Jean, and Hubert Troestler (eds). 1994. *Actes du Colloque sur la problématique de l'aménagement linguistique : enjeux théoriques et pratiques*. Québec: Office de la langue française.
- Doyle, Michael Scott. 2003. "Translation Pedagogy and Assessment: Adopting ATA's Framework for Standard Error Marking." *ATA Chronicle* November/December: 21–29.
- Drozd, Lubomir. 1971. *Anmerkung zum Dokument JAB/DS, 440/71*, Typescript, 2 p.
- Drouin, Patrick. 2004. "Detection of Domain Specific Terminology Using Corpora Comparison." In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, 79–82. Lisbon: European Language Resources Association.
- Drouin, Patrick, Natalia Grabar, Thierry Hamon, Kyo Kageura, and Koichi Takeuchi. 2018. "Computational Terminology and Filtering of Terminological Information." *Special Issue of Terminology* 24 (1): 1–6. <https://doi.org/10.1075/term.00010.dro>
- Drouin, Patrick, Jean-Benoît Morel, and Marie-Claude L'Homme. 2020. "Automatic Term Extraction from Newspaper Corpora: Making the Most of Specificity and Common Features." In *Proceedings of 6th International Workshop on Computational Terminology*, edited by Béatrice Daille, Kyo Kageura, and Ayla Rigouts Terryn, 1–7. Paris: ELRA.
- Dubois, Jean, Mathée Giacomo, Louis Guespin, Jean-Baptiste Marcellesi, Christiane Marcellesi, and Jean-Pierre Mével. 1994. *Dictionnaire de linguistique et des sciences du langage*. Trésors du français. Paris: Larousse.
- Dubuc, Robert. 1975. "Formation des terminologues théoriciens ou praticiens ?" *Banque des mots* 9: 13–22.
- Dubuc, Robert. 1978. *Manuel pratique de terminologie*. Brossard (Quebec): Linguattech.
- Dubuc, Robert. 1985. *Manuel pratique de terminologie*, 2nd ed. Brossard (Quebec): Linguattech.
- Dubuc, Robert. 1992. *Manuel pratique de terminologie*. 3rd ed. Brossard (Quebec): Linguattech.
- Dubuc, Robert. 2002. *Manuel pratique de terminologie*. 4th ed. Brossard (Quebec): Linguattech.
- Ducháček, Otto. 1979. "La synonymie en terminologie." In *Les Actes du 6ème Colloque International de Terminologie*, edited by Office de la Langue Française, 107–118. Quebec: Éditeur Officiel du Québec.
- Duquet-Picard, Diane. 1986. *La synonymie en langues de spécialité: étude du problème en terminologie*. Quebec: GIRSTERM.
- Durán-Muñoz, Isabel, and Marie-Claude L'Homme. 2020. "Diving into English Motion Verbs from a Lexico-Semantic Approach: A Corpus-based Analysis on Adventure Tourism." *Terminology* 26 (1): 34–60. <https://doi.org/10.1075/term.00041.dur>
- Dury, Pascaline. 1999. "Étude comparative et diachronique des concepts ecosystem et écosystème." *Meta* 44 (3): 485–499. <https://doi.org/10.7202/002690ar>
- Dury, Pascaline. 2004. "Building a Bilingual Diachronic Corpus of Ecology: The Long Road to Completion." *ICAME Journal* 28: 5–16.
- Dury, Pascaline. 2012. "Le sentiment d'un besoin néologique chez l'expert pour remplacer un terme à connotation péjorative. Quelques exemples tirés du domaine médical." *Neologica* 6: 81–93.

- Dury, Pascaline. 2013. "Que montre l'étude de la variation d'une terminologie dans le temps. Quelques pistes de réflexion appliquées au domaine médical." *Debate Terminológico* 9: 2–10.
- Dury, Pascaline. 2018. "La 'médicalisation' de la terminologie du textile en anglais, une analyse diachronique basée sur corpus." *Textes et Contextes* 13 (2). (<https://preo.u-bourgogne.fr/textesetcontextes/index.php?id=2312#quotation>). Accessed 7 December 2021.
- Dury, Pascaline, and Patrick Drouin. 2011. "When Terms Disappear from a Specialized Lexicon: A Semi-automatic Investigation into 'Necrology'." *ICAME Journal* 35: 19–33.
- Dury, Pascaline, and Susanne Lervad. 2008. "La variation synonymique dans la terminologie de l'énergie: approches synchronique et diachronique, deux études de cas." *LSP & Professional Communication* 8 (2): 66–78.
- Dury, Pascaline, and Aurélie Picton. 2009. "Terminologie et diachronie : vers une réconciliation théorique et méthodologique?" *Revue française de linguistique appliquée* 14 (2): 31–41. <https://doi.org/10.3917/rfla.142.0031>
- Duvivier-Senis, Angela. 2016. "John Rupert Firth, Historien de la linguistique et fondateur de la 'London School'." PhD diss., Université de Paris 7.

E

- Eck, Karen, and Ingrid Meyer. 1995. "Bringing Aristotle into the 20th Century. Computer-Aided Definition Construction in a Terminological Knowledge Base." In *Standardizing and Harmonizing Terminology: Theory and Practice*, edited by Sue Ellen Wright, and Richard A. Strehlow, 83–100. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/STP137485>
- Eensoo, Egle, and Mathieu Valette. 2013. *Sémantique textuelle et TA : un exemple d'application à l'analyse des sentiments*. Working paper (<https://halshs.archives-ouvertes.fr/halshs-00968634>). Accessed 7 December 2021.
- Egger, Clara, Aurélie Picton, and Doris Schopper. 2018. "How Do Humanitarian IO Conceptualise the World? An Interdisciplinary Approach to the Analysis of Concepts Used in Humanitarian Action." Paper presented at the RUN Workshop, Geneva, Switzerland.
- Eisenstein, Jacob. 2019. *Introduction to Natural Language Processing*. Cambridge, MA: MIT Press.
- El Boukkouri, Hicham, Olivier Ferret, Thomas Lavergne, and Pierre Zweigenbaum. 2019. "Embedding Strategies for Specialized Domains: Application to Clinical Entity Recognition." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL) (Student Research Workshop)*, edited by Fernando Alva-Manchego, Eunsol Choi, and Daniel Khashabi, 295–301. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-2041>
- Elchacar, Mireille. 2016. "Etude diachronique de néologismes du vocabulaire sociopolitique. La vitalité de antimondialisation, altermondialiste et du fractomorphème *alter* – dix ans après leur apparition dans la presse générale." *Neologica* 10: 75–99.
- Ethayarajh, Kavin. 2019. "How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, edited by Sebastian Padó, and Ruihong Huang, 55–65. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/D19-1006>

- Evans, Vyvyan. 2004. *The Structure of Time: Language, Meaning and Temporal Cognition*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/hcp.12>
- Evans, Vyvyan, and Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.
- Evert, Stefan. 2004. "The Statistics of Word Cooccurrences: Word Pairs and Collocations." PhD diss., University of Stuttgart.
- Evert, Stefan, Peter Uhrig, Sabine Bartsch, and Thomas Proisl. 2017. "E-VIEW-alation – A Large-scale Evaluation Study of Association Measures for Collocation Identification." In *Proceedings of eLex 2017 – Electronic lexicography in the 21st century: Lexicography from Scratch*, edited by Iztok Kosem, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek, and Vít Baisa, 531–549. Brno: Lexical Computing CZ s.r.o.

F

- Faber, Pamela. 2002. "Terminographic Definition and Concept Representation." In *Training the Language Services Provider for the New Millennium: Proceedings of the III Encontros de Tradução de Astra-FLUP*, edited by Johann Haller, and Margherita Ulrych, 343–354. Porto: Faculdade de Letras, Universidade do Porto.
- Faber, Pamela. 2009. "The Cognitive Shift in Terminology and Specialized Translation." *MonTi Monografías de Traducción e Interpretación* 1 (1): 107–134. <https://doi.org/10.6035/MonTi.2009.1.5>
- Faber, Pamela. 2011. "The Dynamics of Specialized Knowledge Representation: Simulational Reconstruction or the Perception–Action Interface." *Terminology* 17 (1): 9–29. <https://doi.org/10.1075/term.17.1.02fab>
- Faber, Pamela, (Ed.) 2012. *A Cognitive Linguistics View of Terminology and Specialized Language*. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110277203>
- Faber, Pamela. 2015. "Frames as a Framework for Terminology." In *Handbook of Terminology*, edited by Henrik J. Kockaert, and Frieda Steurs, 14–33. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/hot.1.02fra1>
- Faber, Pamela, and Miriam Buendía Castro. 2014. "EcoLexicon." In *Proceedings of the XVI EU-RALEX International Congress*, edited by Andrea Abel, Chiara Vettori, and Natascia Ralli, 601–607. Bolzano, Italy.
- Faber, Pamela, and Pilar León-Araúz. 2014. "Specialized Knowledge Dynamics: From Cognition to Culture-bound Terminology." In *Dynamics and Terminology. An Interdisciplinary Perspective on Monolingual and Multilingual Culture-bound Communication*, edited by Rita Temmerman, and Marc Van Campenhoudt, 135–159. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.16.08fab>
- Faber, Pamela, and Pilar León-Araúz. 2016. "Specialized Knowledge Representation and the Parameterization of Context." *Frontiers in Psychology* 7: 196. <https://doi.org/10.3389/fpsyg.2016.00196>
- Faber, Pamela, and Pilar León-Araúz. 2019. "Frame-Based Terminology Applied to Military Science: Transforming a Glossary into a Knowledge Resource." *Lexicography* 6: 105–131. <https://doi.org/10.1007/s40607-019-00060-y>
- Faber, Pamela, Pilar León Araúz, and Juan Antonio Prieto Velasco. 2009. "Semantic Relations, Dynamicity, and Terminological Knowledge Bases." *Current Issues in Language Studies* 1 (1): 1–23.

- Faber, Pamela, Pilar León-Araúz, Juan Antonio Prieto Velasco, and Arianne Reimerink. 2007. "Linking Images and Words: The Description of Specialized Concepts." *International Journal of Lexicography* 20 (1): 39–65. <https://doi.org/10.1093/ijl/eclo38>
- Faber, Pamela, Pilar León-Araúz, and Arianne Reimerink. 2011. "Knowledge representation in EcoLexicon." In *Technological Innovation in the Teaching and Processing of LSPs: Proceedings of TISLID'10*, edited by Noa Talaván Zanón, Elena Martínez Monje, and Francisco Palazón Romero, 367–386. Madrid: Universidad Nacional de Educación a Distancia.
- Faber, Pamela, Pilar León-Araúz, and Arianne Reimerink. 2014. "Representing Environmental Knowledge in EcoLexicon." In *Languages for Specific Purposes in the Digital Era*, edited by Elena Bárcena, Timothy Read, and Jorge Arús, 267–301. Educational Linguistics 19. Cham: Springer. https://doi.org/10.1007/978-3-319-02222-2_13
- Faber, Pamela, Pilar León-Araúz, and Arianne Reimerink. 2016. "EcoLexicon: New Features and Challenges." In *GLOBALEX 2016: Lexicographic Resources for Human Language Technology in Conjunction with the 10th Edition of the Language Resources and Evaluation Conference*, edited by Ilan Kernerman, Iztok Kosem, Simon Krek, and Lars Trap-Jensen, 73–80. Portorož.
- Faber, Pamela, Ricardo Mairal, and Pedro Magaña. 2011. "Linking a Domain-Specific Ontology to a General Ontology." In *Proceedings of the Twenty-Fourth International Florida Artificial Intelligence Research Society Conference*, edited by Charles R. Murray, and Philip M. McCarthy, 564–569. Menlo Park, CA: AAAI Press.
- Faber, Pamela, and Laura Medina-Rull. 2017. "Written in the Wind: Cultural Variation in Terminology." In *Cognitive Approaches to Specialist Languages*, edited by Marcin Grygiel, 419–442. Newcastle-upon-Tyne: Cambridge Scholars.
- Faber, Pamela, Silvia Montero Martínez, María Rosa Castro Prieto, José Senso Ruiz, Juan Antonio Prieto Valesco, Pilar León Araúz, Carlos Márques Linares, and Miguel Vega Expósito. 2006. "Process-oriented Terminology Management in the Domain of Coastal Engineering." *Terminology* 12 (2): 189–213. <https://doi.org/10.1075/term.12.2.o3fab>
- Faber, Pamela, Juan Verdejo, Pilar León-Araúz, Arianne Reimerink, and Gloria Guzman. 2014. "Neural Substrates of Specialized Knowledge Representation: an fMRI Study." *Revue française de linguistique appliquée* 19: 15–34. <https://doi.org/10.3917/rfla.191.0015>
- Fabre, Cécile. 2015. "Sémantique distributionnelle automatique : la proximité distributionnelle comme accès au sens." *Études de linguistique appliquée (Linguistique et informatique)* 4: 395–405. <https://doi.org/10.3917/ela.180.o395>
- Fabre, Cécile, Nabil Hathout, Franck Sajous, and Ludovic Tanguy. 2014. "Ajuster l'analyse distributionnelle à un corpus spécialisé de petite taille." In *Actes de la 21^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2014) Atelier Sémantique Distributionnelle (SemDis 2014)*, edited by Cécile Fabre, Nabil Hathout, Lydia-Mai Ho-Dac, François Morlane-Hondère, Philippe Muller, Franck Sajous, Ludovic Tanguy, and Tim van de Cruys, 266–279. Marseille, France.
- Fabre, Cécile, and Alessandro Lenci. 2015. "Distributional Semantics Today: Introduction to the special issue." *Traitement Automatique des Langues* 56 (2): 7–20.
- Fan, Yadan, Serguei Pakhomov, Reed McEwan, Wendi Zhao, Elizabeth Lindemann, and Rui Zhang. 2019. "Using Word Embeddings to Expand Terminology of Dietary Supplements on Clinical Notes." *JAMIA Open* 2 (2): 246–253. <https://doi.org/10.1093/jamiaopen/ooz007>
- Faruqui, Manaal, Hinrich Schütze, Isabel Trancoso, and Yadollah Yaghoobzadeh (eds). 2017. *Proceedings of the First Workshop on Subword and Character Level Models in NLP*. Stroudsburg, PA: Association for Computational Linguistics.

- Fauconnier, Gilles, and Mark Turner. 1996. "Blending as a Central Process of Grammar." In *Conceptual Structure, Discourse, and Language*, edited by Adele Goldberg, 113–130. Stanford: CSLI Publications.
- Fauconnier, Gilles, and Mark Turner. 1998. "Conceptual Integration Networks." *Cognitive Science* 22 (2): 133–187. https://doi.org/10.1207/s15516709cog2202_1
- Fauconnier, Gilles, and Mark Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.
- Faulstich, Enilde. 1995. "Socioterminologia: mais que um método de pesquisa, uma disciplina." *Ciência da Informação* 24 (3). (<http://revista.ibict.br/ciinf/article/view/566>). Accessed 7 December 2021.
- Faulstich, Enilde. 1998. "Variação terminológica: algumas tendências no português do Brasil." In *Lèxic, corpus i diccionaris. Cicle de conferències 96–97*, edited by Maria Teresa Cabré, 141–154. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Fedor de Diego, Alicia. 1995. *Terminología: teoría y práctica*. Caracas: Equinoccio.
- Fedor de Diego, Alicia, and Michele Boroni de Sánchez-Vegas (eds). 1984. *Actas del primer seminario nacional de terminología*. Caracas: Universidad Simon Bolívar y Unión Latina.
- Felber, Helmut. 1980. "International Standardization of Terminology: Theoretical and Methodological Aspects." *International Journal of the Sociology of Language* 23: 65–79. <https://doi.org/10.1515/ijsl.1980.23.65>
- Felber, Helmut. 1981a. "The Vienna School of Terminology Fundamentals and its Theory." In *Theoretical and Methodological Problems of Terminology: International Symposium Proceedings*, edited by Magdalena Krommer-Benz, 69–86. Vol 6 of Infoterm Series. Munich, London, Paris: Sauer.
- Felber, Helmut. 1981b. *Nachtrag: Änderungen seit 1971* [to Wüster's 1971 Grundsätze der fachsprachlichen Normung]. In *Fachsprachen*, edited by Walther von Hahn, 273–275. Darmstadt, Wissenschaftliche Buchgesellschaft.
- Felber, Helmut. 1984. *Terminology Manual*. Paris: UNESCO (Infoterm).
- Felber, Helmut. 1987. *Manuel de terminologie*. Paris: UNESCO (Infoterm).
- Felber, Helmut. 1993 [1986]. "Einige Grundfragen der Terminologie-Wissenschaft aus der Sicht der Allgemeinen Terminologielehre." In *Ausgewählte Texte zur Terminologie*, edited by Christer Laurén, and Heribert Picht, 377–396. Vienna, TermNet, Infoterm.
- Felber, Helmut. 1998. Historische Aspekte der Entwicklung der Allgemeinen Terminologielehre bei Wüster, Aug. –Oct. 1998, typescript.
- Felber, Helmut, Friedrich Lang, and Gernot Wersig (eds). 1979. *Terminologie als angewandte Sprachwissenschaft. Gedenkschrift für Univ-Prof. Dr. Eugen Wüster*. München: Sauer.
- Felber, Helmut, and Heribert Picht. 1984. *Métodos de terminografía y principios de investigación terminológica*. Madrid: Instituto Cervantes.
- Feliu, Judit. 2004. "Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica." PhD diss., Universitat Pompeu Fabra.
- Ferella, Chiara. 2020. "Early Greek Medical Metaphors and the Question of Deliberateness." In *Drawing Attention to Metaphor: Case Studies across Time Periods, Cultures and Modalities*, edited by Camilla di Biase-Dyson, and Markus Egg, 129–158. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/ftl.5.06fer>
- Fernández, Trinidad, and Pamela Faber. 2011. "The Representation of Multidimensionality in a Bilingualized English-Spanish Thesaurus for Learners in Architecture and Building Construction." *International Journal of Lexicography* 24 (2): 198–225. <https://doi.org/10.1093/ijl/ecq046>

- Fernández-López, Mariano, Asunción Gómez-Pérez, and Natalia Juristo. 1997. "METHONTOL-OGY: From Ontological Art Towards Ontological Engineering." In *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*, 33–40. Stanford University, Stanford.
- Fernández-López, Mariano, María Poveda-Villalón, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 2019. "Why are Ontologies not Reused across the Same Domain?" In *Journal of Web Semantics* 57: 100492–100501. <https://doi.org/10.1016/j.websem.2018.12.010>
- Fernández-Silva, Sabela. 2011. "Variación terminológica y cognición: Factores cognitivos en la denominación del concepto especializado." PhD diss., Universitat Pompeu Fabra.
- Fernández-Silva, Sabela. 2013. "Variación denominativa y punto de vista." *Revista Debate Terminológico* 9: 11–37.
- Fernández-Silva, Sabela. 2014. "El poder retórico de la variación terminológica." *Debate Terminológico*, 12: 55–59.
- Fernández-Silva, Sabela. 2016. "The Cognitive and Rhetorical Role of Term Variation and its Contribution to Knowledge Construction in Research Articles." *Terminology* 22 (1): 52–79. <https://doi.org/10.1075/term.22.1.03fer>
- Fernández-Silva, Sabela. 2018. "A investigación sobre variación terminológica: avances actuales e desafíos metodológicos." In *Terminología: a necesidade da colaboración*, edited by Manuel González, María-Dolores Sánchez-Palomino, and Inés Veiga, 39–66. Madrid: Iberoamericana / Vervuert.
- Fernández-Silva, Sabela. 2019. "The Cognitive and Communicative Functions of Term Variation in Research Articles: A Comparative Study in Psychology and Geology." *Applied Linguistics* 40 (4): 624–645. <https://doi.org/10.1093/applin/amy004>
- Fernández-Silva, Sabela, Judit Freixa, and María Teresa Cabré. 2014. "A method for analysing the dynamics of naming from a monolingual and multilingual perspective." In *Dynamics and Terminology. An Interdisciplinary Perspective on Monolingual and Multilingual Culture-bound Communication*, edited by Rita Temmerman, and Marc Van Campenhout, 183–211. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.16.10fer>
- Fernández-Silva, Sabela, and Maribel Tercedor Sánchez. 2019. "El desarrollo de la competencia sobre variación terminológica en estudiantes de traducción e interpretación: una experiencia de colaboración interuniversitaria." *RLA, Revista de lingüística teórica y aplicada* 57 (2): 137–163. <https://doi.org/10.4067/S0718-48832019000200137>
- Ferrari, Laura. 2002. "¿Un caso de polisemia en el discurso jurídico?" *Terminology* 8 (2): 221–244. <https://doi.org/10.1075/term.8.2.04fer>
- Ferret, Olivier. 2014. "Utiliser un modèle neuronal générique pour la substitution lexicale." In *Actes de la 21^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2014) Atelier Sémantique Distributionnelle (SemDis 2014)*, edited by Cécile Fabre, Nabil Hathout, Lydia-Mai Ho-Dac, François Morlane-Hondère, Philippe Muller, Franck Sajous, Ludovic Tanguy, and Tim van de Cruys, 218–227. Marseille, France.
- Fillmore, Charles. 1968. "The Case for Case." In *Universals in Linguistics Theory*, edited by Emmon Bach, and Robert T. Harms, 1–88. New York: Holt, Rinehart and Winston.
- Fillmore, Charles. 1982. "Frame Semantics." In *Linguistics in the Morning Calm*, edited by the Linguistic Society of Korea, 111–138. Seoul: Hanshin.
- Fillmore, Charles. 1985. "Frames and the Semantics of Understanding." *Quaderni di Semantica* 6 (2): 222–254.
- Fillmore, Charles, and Collin Baker. 2010. "A Frames Approach to Semantic Analysis." In *The Oxford Handbook of Linguistic Analysis*, edited by Bernard Heine, and Heiko Narrog, 313–339. Oxford: Oxford University Press.

- Firth, John Rupert. 1957. "Modes of Meaning." In *Papers in Linguistics* 1934–1951, edited by John R. Firth, 190–215. London: Oxford University Press.
- Firth, John Rupert. 1968 [1952]. "Linguistic Analysis as a Study of Meaning." In *Selected Papers of J.R. Firth*, edited by Frank Robert Palmer, 12–26. London: Longman.
- Fløttum, Kjersti, Trine Dahl, and Torodd Kinn. 2006. *Academic Voices: Across Languages and Disciplines*. Pragmatics & Beyond New Series 148. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.148>
- Foucault, Michel. 1968. "Sur l'archéologie des sciences: réponse au cercle d'épistémologie." *Cahiers pour l'Analyse* 9 (Généalogie des Sciences): 9–40.
- Foucault, Michel. 1969. *L'Archéologie du savoir*. Paris: Gallimard.
- Frantzi, Katerina T., and Sophia Ananiadou. 1999. "The C-value/NC-value Domain-Independent Method for Multi-Word Term Extraction." *Journal of Natural Language Processing* 6 (3): 145–179. https://doi.org/10.5715/jnlp.6.3_145
- Frawley, William. 1998. "New Forms of Specialized Dictionaries." *International Journal of Lexicography* 1 (3): 89–213.
- Freixa, Judit. 2002. "La variació terminològica: Anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient." PhD diss., Universitat Pompeu Fabra.
- Freixa, Judit. 2005. "Variación terminológica: ¿Por qué y para qué?" *Meta* 50 (4). <https://doi.org/10.7202/019917ar>
- Freixa, Judit. 2006. "Causes of Denominative Variation in Terminology: A Typology Proposal." *Terminology* 12 (1): 51–77. <https://doi.org/10.1075/term.12.1.04fre>
- Freixa, Judit. 2013. "Otra vez sobre las causas de la variación denominativa." *Debate Terminológico* 9: 38–46.
- Freixa, Judit. 2019. "Augment constant de manlleus en la premsa catalana: una conseqüència de la mundialització?" *Treballs de Sociolingüística Catalana* 29: 43–64.
- Freixa, Judit, and Sabela Fernández-Silva. 2017. "Terminological Variation and the Unsaturability of Concepts." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francœur, John Humbley, and Aurélie Picton, 155–181. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.07fre>
- Freixa, Judit, and Amor Montané. 2006. "Variación denominativa y biunivocidad en el lenguaje de las matemáticas." *Revista Española de Lingüística* 36 (1): 189–215.
- Frérot, Cécile. 2018. "Enseignement de la terminologie appliquée à une formation universitaire professionnalisante : illustration d'une collaboration avec l'Organisation Mondiale de la Propriété intellectuelle." *Myriades* 4: 33–52.
- Fries, Marie-Hélène. 2016. *Nanomonde et nouveau monde. Quelques métaphores clés sur les nanotechnologies aux Etats-Unis*. Grenoble: ELLUG.
- Fuentes-Arderiu, Xavier. 2006. "Contra la sinonimia y la polisemia en los lenguajes de especialidad." *Panace@* 7 (24): 241.

G

- Gaballo, Viviana. 2010. "Translating Stones: A Corpus-based Linguistic and Lexicographic Study in Specialized Terminology." *Lexis, Journal of English Lexicology* 4: 55–63. <https://doi.org/10.4000/lexis.583>
- Gagné, Anne-Marie, and Marie-Claude L'Homme. 2016. "Opposite Relationships in Terminology." *Terminology* 22 (1): 30–51. <https://doi.org/10.1075/term.22.1.02gag>

- Gaines, B. 1990. "Knowledge Acquisition Systems" In *Knowledge Engineering. Vol. I: Fundamentals*, edited by Hojjat Adeli, 52–99. New York: McGraw-Hill.
- Galinski, Christian. 2019. "Blütezeit der Zusammenarbeit zwischen Terminologie einerseits und Information und Dokumentation." In *Terminologie: Epochen – Schwerpunkte – Umsetzungen: Zum 25-jährigen Bestehen des Rats für deutschsprachige Terminologie*, edited by Petra Drewer, and Donatella Pulitano, 21–43. Wiesbaden: Springer Vieweg.
https://doi.org/10.1007/978-3-662-58949-6_3
- Galinski, Christian, and Jürgen W. Goebel. 1996. *Guide to Terminology Agreements*. Paris/Vienna: European Language Resources Association/ TermNet. (http://www.infoterm.info/pdf/activities/Guide_to_terminology_agreements/GuidetoTerminologyAgreements_FV_2011-02-12_DE.pdf). Accessed 7 December 2021.
- Gallardo, Susana. 2000. "Clon y clonación: Variabilidad conceptual en ciencias biológicas." *Terminology* 6 (1): 59–78. <https://doi.org/10.1075/term.6.1.04gal>
- Gallese, Victor, and George Lakoff. 2005. "The Brain's Concepts: The Role of the Sensory-Motor System in Conceptual Knowledge." *Cognitive Neuropsychology* 22 (3/4): 455–479.
<https://doi.org/10.1080/02643290442000310>
- Gambier, Yves. 1987. "Problèmes terminologiques des pluies acides : pour une socioterminologie." *Meta* 32 (3): 314–320. <https://doi.org/10.7202/002791ar>
- Gambier, Yves. 1991a. "Présupposés de la terminologie: vers une remise en cause." *Cahiers de linguistique sociale* 18: 31–58.
- Gambier, Yves. 1991b. "Travail et vocabulaire spécialisés : prolégomènes à une socio-terminologie." *Meta* 36 (1): 8–15. <https://doi.org/10.7202/002795ar>
- Gambier, Yves. 1998. "Le français dans les communications spécialisées : bilan mitigé." In *Discours professionnels en français*, edited by Yves Gambier, 35–62. Bern: Peter Lang.
- Gambier, Yves, and François Gaudin (eds). 1993. *Socioterminologie: Le Langage et l'homme* 28 (4).
- Gamper, Johann, and Oliviero Stock. 1998. "Corpus-based Terminology." *Terminology* 5 (2): 147–160. <https://doi.org/10.1075/term.5.2.05gam>
- Gangemi, Aldo. 2005. "Ontology Design Patterns for Semantic Web Content." In *Proceedings of International Semantic Web Conference (ISWC05)*, edited by Yolanda Gil, Enrico Motta, Richard Benjamins, and Mark Musen, 262–276. Berlin/Heidelberg: Springer.
https://doi.org/10.1007/11574620_21
- Gangemi, Aldo, Domenico M. Pisanelli, and Geri Steve. 1998. "Ontology Integration: Experiences with Medical Terminologies." In *Formal Ontology in Information Systems*, edited by Nicola Guarino, 163–178. Amsterdam: IOS Press
- Garcés-Mascareñas, Blanca, and Rinus Penninx. 2016. "Introduction: Integration as a Three-Way Process Approach?" In *Integration Processes and Policies in Europe*, edited by Blanca Garcés-Mascareñas, and Rinus Penninx, 1–9. IMISCOE Research Series. Springer.
https://doi.org/10.1007/978-3-319-21674-4_1
- Garcia, Daniela. 1997. "Structuration du lexique de la causalité et réalisation d'un outil d'aide au repérage de l'action dans les textes." In *Actes des deuxièmes rencontres – Terminologie et Intelligence Artificielle, TIA '97*, 7–26. Toulouse, France.
- Garcia, Marcos, Marcos García-Salido, and Margarita Ramos. 2019. "A Comparison of Statistical Association Measures for Identifying Dependency-based Collocations in Various Languages." In *Proceedings of the Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)*, edited by Agata Savary, Carla Parra Escartín, Francis Bond, Jelena Mitrović, and Verginica Barbu Mititelu, 49–59. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/W19-5107>

- García-Silva, Andrés, Asunción Gómez-Pérez, Mari Carmen Suárez-Figueroa, and Boris Villazón-Terrazas. 2008. "A Pattern Based Approach for Reengineering Non-Ontological Resources into Ontologies." In *3rd Asian Semantic Web Conference ASWC08*, 167–181. Bangkok, Thailand.
- Gaudin, François. 1989. "Terminologie et sociolinguistique." In *Hannoversche Beiträge Zu Sprache Und Kultur. Akten Des Deutsch-Französischen Kolloquiums in Hannover Vom 1: 46–54*. Hannover: University of Hannover.
- Gaudin, François. 1991. "La socioterminologie : un champ en voie de constitution." In *Sociolinguistique, didactique du F.L.E. (deux domaines en mutation)*, *Cahiers de linguistique sociale*, 173–183. Rouen: Université de Rouen.
- Gaudin, François. 1993. *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*. Rouen: Université de Rouen.
- Gaudin, François. 1994. "L'insécurité linguistique des scientifiques. A propos d'une enquête socioterminologique." *Cahiers de lexicologie* 65 (2): 47–58.
- Gaudin, François (ed.). 1995. Usages sociaux des termes: théories et terrains. *Meta* 40 (2).
- Gaudin, François. 1996. "Terminologie: l'ombre du concept." *Meta* 41 (4): 604–621.
<https://doi.org/10.7202/002813ar>
- Gaudin, François. 2003. *Socioterminologie: Une approche sociolinguistique de la terminologie*. Louvain-la-Neuve: De Boeck Supérieur. <https://doi.org/10.3917/dbu.gaudi.2003.01>
- Gaudin, François. 2005. "Socioterminologie." *Langages* 157: 80–92.
<https://doi.org/10.3406/lgge.2005.976>
- Gaudin, François, and Allal Assal (eds). 1991. *Terminologie et sociolinguistique, Cahiers de linguistique sociale*. Rouen: Université de Rouen.
- Gaudin, François, and Myriam Bouveret. 1997. "Partage des noms, partage des notions ? Approche sociolinguistique de difficultés terminologiques en situation interdisciplinaire." In *Autour de la dénomination*, edited by Claude Boisson, and Philippe Thoirion, 241–67. Lyon: Presses universitaires de Lyon.
- Gaudin, François, and Valérie Delavigne. 1995. "L'enquête en terminologie: point de la question et propositions In *Actes des Journées d'études: l'implantation terminologique*, 37–42. Toulouse, France.
- Gaudin, François, and Cristina Nicolae. 2016. "La référencement en socioterminologie: réflexions à partir de l'astronomie." *Repères DoRiF* 10. <https://www.dorif.it/reperes/francois-gaudin-et-cristina-nicolae-la-referenciation-en-socioterminologie-reflexions-a-partir-de-lastronomie/>. Accessed 7 December 2021.
- Gaussier, Éric. 2001. "General Considerations on Bilingual Terminology Extraction." In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, 167–182. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/nlp.2.09gau>
- Geeraerts, Dirk. 1988. "Where does Prototypicality Come from?" In *Topics in Cognitive Linguistics*, edited by Brygida Rudzka-Ostyn, 207–229. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/cilt.50.09gee>
- Geeraerts, Dirk. 2015. "Forward." In *Handbook of Terminology*, edited by Henrik Kockaert, and Frieda Steurs, xvii–xix. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/hot.1.002for>
- Gentner, Dedre, and Brian Bowdle. 2001. "Convention, Form, and Figurative Language Processing." *Metaphor and Symbol* 16: 223–247. <https://doi.org/10.1080/10926488.2001.9678896>

- Gentner, Dedre, and Mary Jo Rattermann. 1991. "Language and the Career of Similarity." In *Similarity and Analogical Reasoning*, edited by Susan Gelman, and James Brynes, 199–241. Cambridge: Cambridge University Press.
- Gerzymisch-Arbogast, Heidrun. 1996. *Termini im Kontext: Verfahren zur Erschließung und Übersetzung der textspezifischen Bedeutung von fachlichen Ausdrücken*. Tübingen: Gunter Narr.
- Ghaddar, Abbas and Philippe Langlais. 2020. "SEDAR: a Large Scale French-English Financial Domain Parallel Corpus." In *Proceedings of the 12th Language Resources and Evaluation, LREC 2020*, edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, et al., 3595–3602. Paris: ELRA.
- Gibbs, Raymond. 2011. "Evaluating Conceptual Metaphor Theory." *Discourse Processes* 48 (8): 529–562. <https://doi.org/10.1080/0163853X.2011.606103>
- Gibbs, Raymond. 2015. "Does Deliberate Metaphor Theory have a Future?" *Journal of Pragmatics* 90: 73–76. <https://doi.org/10.1016/j.pragma.2015.03.016>
- Gil-Berrozpe, Juan Carlos, Pilar León-Araúz, and Pamela Faber. 2016. "Subtypes of Hyponymy in the Environmental Domain: Entities and Processes." In *Terminologie et Ontologie: Théories et Applications. Actes de la conférence TOTH 2016 Chambéry – 9 & 10 juin 2016*, edited by Christophe Roche, 39–54. Chambéry: Éditions de l'Université Savoie Mont Blanc.
- Gil-Berrozpe, Juan Carlos, Pilar León-Araúz, and Pamela Faber. 2019. "Ontological Knowledge Enhancement in EcoLexicon." In *Proceedings of the eLex 2019 Conference: Electronic Lexicography in the 21st century*, edited by Iztok Kosem, Tanara Zingano-Kuhn, Margarita Correia, José Pedro Ferreira, Maarten Jansen, Isabel Pereira, Jelena Kallas, Milos Jakubiček, Simon Krek, and Carole Tiberius, 177–197. Brno: Lexical Computing CZ, s.r.o.
- Gilbert, Pierre. 1977. "Comportement des néologismes scientifiques et techniques dans le temps." In *Actes du 6^{ème} colloque international de terminologie*, edited by Rosita Harvey, and Lise Lebel, 65–74. Québec: Office de la langue française.
- Gillam, Lee, Mariam Tariq, and Khurshid Ahmad. 2005. "Terminology and the Construction of Ontology." *Terminology* 11 (1): 55–81. <https://doi.org/10.1075/term.11.1.04gil>
- Girju, Roxana, Adriana Badulescu, and Dan Moldovan. 2003. "Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relation." In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 80–87. Sapporo, Japan: ACL. <https://doi.org/10.3115/1073445.1073456>
- Gledhill, Christopher. 2000. *Collocations in Science Writing*. Language in Performance Series no. 22. Tübingen: Gunter Narr Verlag.
- Goffin, Roger. 1989. "La terminologie des sciences et des techniques nucléaires. Un cas de diachronie récente." In *Terminologie diachronique, Proceedings of the First International Conference on Diachronic Terminology*, edited by Caroline De Schaetzen, 94–107. Paris: CILF.
- Goffin, Roger. 1997. "Eurodicautom, la banque de données terminologiques de la commission européenne (1973–1996)." *Terminologie et Traduction* 2: 29–73.
- Goldberg, Adele. 1998. "Patterns of Experience in Patterns of Language." In *The New Psychology of Language*, edited by Michael Tomasello, 203–219. New York: Routledge.
- Goldstone, Robert L., Alan Kersten, Paulo F. Carvalho. 2012. "Concepts and Categorization." In *Handbook of Psychology: Experimental Psychology*, edited by Alice F. Healy, Robert W. Proctor, and Irving B. Weiner, 607–630. Hoboken, NJ: John Wiley & Sons, Inc.
- Gómez-Pérez, Asunción, Mariano Fernández-López, and Óscar Corcho. 2003. *Ontological Engineering*. Berlin/Heidelberg: Springer.
- Gómez-Pérez, Asunción, and David Manzano-Macho. 2004. "An Overview of Methods and Tools for Ontology Learning from Texts." In *The Knowledge Engineering Review* 19 (3), 187–212. <https://doi.org/10.1017/S0269888905000251>

- Good, Irvin John. 1953. "The Population Frequencies of Species and the Estimation of Population Parameters" *Biometrika* 40 (3–4): 237–264. <https://doi.org/10.1093/biomet/40.3-4.237>
- Good, Irvin John and George H. Toulmin. 1956. "The Number of New Species, and the Increase in Population Coverage, when a Sample is Increased." *Biometrika* 43 (1–2): 45–63. <https://doi.org/10.1093/biomet/43.1-2.45>
- Gotti, Maurizio, and Davide Simone Giannoni (eds). 2006. *New Trends in Specialised Discourse Analysis*. Bern: Peter Lang.
- Grabar, Natalie, Cyril Grouin, Thierry Hamon, and Vincent Claveau. 2019. "Annotated Corpus with Clinical Cases in French." In *Actes de la 26th conférence sur le Traitement Automatique des Langues Naturelles (TALN'2019)*, 1–14. Toulouse, France.
- Grady, Joseph. 1997a. "Foundations of Meaning: Primary Metaphors and Primary scenes." PhD diss., Berkeley: University of California.
- Grady, Joseph. 1997b. "Theories and Buildings Revisited." *Cognitive Linguistics* 8 (4): 267–290. <https://doi.org/10.1515/cogl.1997.8.4.267>
- Grady, Joseph. 1999. "A Typology of Motivation for Conceptual Metaphor: Correlation vs. Resemblance." In *Metaphor in Cognitive Linguistics*, edited by Raymond W. Gibbs, and Gerard J. Steen, 79–100. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/cilt.175.06gra>
- Grady, Joseph. 2005. "Primary Metaphors as Inputs to Conceptual Integration." *Journal of Pragmatics* 37 (10): 1595–1614. <https://doi.org/10.1016/j.pragma.2004.03.012>
- Grady, Joseph, and Christopher Johnson. 2002. "Converging Evidence for the Notions of 'Subscene' and 'Primary Scene'." In *Metaphor and Metonymy in Comparison and Contrast*, edited René Dirven, and Ralf Pörings, 533–554. Berlin/New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110219197.533>
- Gregory, Michael, and Susanne Carroll. 1978. *Language and Situation: Language Varieties and their Social Contexts*. London, Henley and Boston: Routledge & Kegan Paul.
- Grön, Leonie, and Ann Bertels. 2018. "Clinical Sublanguages: Vocabulary Structure and Its Impact on Term Weighting." *Terminology* 24 (1): 41–65. <https://doi.org/10.1075/term.00013.gro>
- Gross, Gaston. 1994. "Classes d'objets et description des verbes." *Langages* 115: 15–30. <https://doi.org/10.3406/lgge.1994.1684>
- Gruber, Thomas Robert. 1993. "A Translation Approach to Portable Ontology Specifications." *Acquis* 5 (2): 199–220. <https://doi.org/10.1006/knac.1993.1008>
- Grüninger, Michael, and Mark S. Fox. 1994. "The Role of Competency Questions in Enterprise Engineering." In *Proceedings of the IFIP Workshop on Benchmarking – Theory and Practice*, 22–31. Trondheim, Norway.
- Grüninger, Michael, and Mark S. Fox. 1995. "Methodology for the Design and Evaluation of Ontologies." In *Proceedings of the IJCAI Workshop on Basic Ontological Issues in Knowledge Sharing*. Seattle, USA. (<http://stl.mie.utoronto.ca/publications/gruninger-ijcai95.pdf>). Accessed 7 December 2021.
- Guarino, Nicola, Daniel Oberle, and Steffen Staab. 2009. "What is an Ontology." In *Handbook on Ontologies*, edited by Steffen Staab, and Rudi Studer, 1–17. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-540-92673-3_0
- Guespin, Louis. 1985. "Matériaux pour une glottopolitique." *Cahiers de linguistique sociale* 7: 13–33.
- Guespin, Louis. 1991. "La circulation terminologique et les rapports entre science, technique, production." *Cahiers de linguistique sociale* 18: 59–80.
- Guespin, Louis. 1993. "Avant-propos." In *Pour une socioterminologie*, edited by François Gaudin, 9–10. Rouen: Publications de l'université de Rouen.

- Guespin, Louis, and Jean-Baptiste Marcellesi. 1986. "Pour la glottopolitique." *Langages* 83: 5–34. <https://doi.org/10.3406/lgge.1986.2493>
- Guilbert, Louis. 1973. "La spécificité du terme scientifique et technique." *Langue française* 17: 5–17. <https://doi.org/10.3406/lfr.1973.5617>

H

- Hahn, Ulrike, and Nick Chater. 1997. "Concepts and Similarity." In *Knowledge, Concepts, and Categories*, edited by Koen Lamberts, and David Shanks, 93–131. Cambridge, MA: MIT Press.
- Hall, David, Daniel Jurafsky, and Christopher Manning. 2008. "Studying the History of Ideas Using Topic Models." In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*, 363–371. Honolulu, Hawaii. <https://doi.org/10.3115/1613715.1613763>
- Halliday, Michael A. K. 1985. "Part A." In *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*, edited by Michael A. K. Halliday, and Ruqaiya Hasan, 3–49. Geelong: Deakin University Press.
- Halliday, Michael A. K. 1999. "The Grammatical Construction of Scientific Knowledge: The Framing of the English Clause." In *The Language of Science*, edited by Jonathan Webster, 102–134. London/New York: Continuum.
- Halliday, Michael A. K. 2004. *The Language of Science*. London/New York: Continuum.
- Halskov, Jakob, and Caroline Barrière. 2008. "Web-based Extraction of Semantic Relation Instances for Terminology Work." *Terminology* 14 (1): 20–44. <https://doi.org/10.1075/term.14.1.03hal>
- Hamed, Fawzi. 2016. "Terminology Variation and Foreign Language Hegemony: The Case of the Arabic-speaking North African Countries." *Translation Journal* July 2016. (<https://translationjournal.net/July-2016/terminology-variation-and-foreign-language-hegemony-the-case-of-the-arabic-speaking-north-african-countries.html>). Accessed 15 December 2020.
- Hamilton, William, Jure Leskovec, and Dan Jurafsky. 2016. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." Last modified October 25, 2018. (<https://arxiv.org/abs/1605.09096>). Accessed 7 December 2021. <https://doi.org/10.18653/v1/P16-1141>
- Hanks, Patrick. 2004. "Corpus Pattern Analysis." In *Proceedings of the XI EURALEX*, edited by Geoffrey Williams, and Sandra Vessier, 87–97. Lorient: Université de Bretagne-Sud.
- Harispe, Sébastien, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. 2015. *Semantic Similarity from Natural Language and Ontology Analysis*. Synthesis Lectures on Human Language Technologies. arXiv:1704.05295 [cs.AI]. San Rafael, CA: Morgan & Claypool. <https://doi.org/10.2200/Soo639ED1Vo1Y201504HLT027>
- Harris, Zellig S. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.
- Harris, Zellig. 1954. "Distributional structure." *Word* 10: 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Harris, Zellig. 1968. *Mathematical Structures of Language*. New York: John Wiley and Sons.
- Hart, Christopher. 2008. "Critical Discourse Analysis and Metaphor: Toward a Theoretical Framework." *Critical Discourse Studies* 5 (2): 91–106. <https://doi.org/10.1080/17405900801990058>
- Hartmann, Reinhard R. K. 2007. "Contrastive Text Analysis and the Search for Equivalence in the Bilingual Dictionary." In *Symposium on Lexicography II*, edited by Karl Hyldgaard-Jensen, and Arne Zettersten, 121–132. Tübingen: Max Niemeyer.
- Hatim, Basil. 1997. *Communication across Cultures. Translation Theory and Contrastive Text Linguistics*. Exeter: University of Exeter Press.

- Hayes-Roth, Frederick. 1984. "The Knowledge-Based Expert System: A Tutorial." *Computer*, September 1984: 11–28. <https://doi.org/10.1109/MC.1984.1659242>
- Hazem, Amir, and Béatrice Daille. 2015. "Méthode semi-compositionnelle pour l'extraction de synonymes des termes complexes." *Traitement automatique des langues* 56 (2): 51–76.
- Hearst, Marti. 1992. "Automatic Acquisition of Hyponyms from Large Text Corpora." In *Proceedings of the 14th International Conference on Computational Linguistics (COLING '92)*, 539–545. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/992133.992154>
- Heath, Tom, and Christian Bizer. 2011. *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology. San Rafael, CA: Morgan & Claypool. <https://doi.org/10.2200/Soo0334ED1Vo1Y201102WBE001>
- Hendrickx, Iris, Zornitsa Kozareva, Preslav Nakov, Diarmuid Ó Séaghdha, Stan Szpakowicz, and Tony Veale. 2013. "SemEval–2013 Task 4: Free Paraphrases of Noun Compounds." In *Second Joint Conference on Lexical and Computational Semantics ("SEM): Proceedings of the Seventh International Workshop on Semantic Evaluation*, Vol. 2, edited by Suresh Manandhar, and Deniz Yuret, 138–143. Atlanta, GA: ACL.
- Hepp, Martin. 2008. "GoodRelations: An Ontology for Describing Products and Services Offers on the Web." In *Proceedings of the 16th international conference on Knowledge Engineering: Practice and Patterns (EKAW '08)*, 329–346. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-540-87696-0_29
- Hesse, Mary. 1966. *Model and Analogies in Science*. Notre Dame, IN: Notre Dame University Press.
- Heylen, Kris, and Ann Bertels. 2016. "Sémantique distributionnelle en linguistique de corpus." *Langages* 201: 51–64. <https://doi.org/10.3917/lang.201.0051>
- Heylen, Kris, Dirk Speelman, and Dirk Geeraerts. 2012. "Looking at Word Meaning. An Interactive Visualization of Semantic Vector Spaces for Dutch Synsets." In *Proceedings of the European Chapter of the Association for Computational Linguistics (EACL 2012): Joint Workshop of LINGVIS et UNCLH*, edited by Miriam Butt, Sheelagh Carpendale, Gerald Penn, Jelena Prokić, and Michael Cysouw, 16–24. Stroudsburg, PA: Association for Computational Linguistics.
- Heylen, Kris, Thomas Wielfaert, Dirk Speelman, and Dirk Geeraerts. 2015. "Monitoring Polysmy: Word Space Models as a Tool for Large-scale Lexical Semantic Analysis." *Lingua* 157: 153–172. <https://doi.org/10.1016/j.lingua.2014.12.001>
- Hilpert, Martin. 2011. "Dynamic Visualizations of Language Change: Motion Charts on the Basis of Bivariate and Multivariate Data from Diachronic Corpora." *International Journal of Corpus Linguistics* 16 (4): 435–461. <https://doi.org/10.1075/ijcl.16.4.01hil>
- Hilpert, Martin, and David Correia Saavedra. 2017. "Using Token-based Semantic Vector Spaces for Corpus-linguistic Analyses: From Practical Applications to Tests of Theoretical Claims." *Corpus Linguistics and Linguistic Theory* 13: 1–32. <https://doi.org/10.1515/cllt-2017-0009>
- Hoffman, Robert. 1980. "Metaphor in Science." In *Cognition and Figurative Language*, edited by Richard Honeck, and Robert Hoffman, 393–423. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Holzem, Maryvonne. 1999. *Terminologie et documentation, pour une meilleure circulation des savoirs*. Paris: ADBS Éditions.
- Holzem, Maryvonne. 2019. "Études de pratiques interprétatives de documents numériques: Apports des sciences de la culture." Habilitation à Diriger des Recherches, Université de Rouen Normandie.

- Holzem, Maryvonne, and Jacques Labiche. 2003. *Dessillement numérique : Énaction, interprétation, connaissances*. Coll. Gramm-R.: études de linguistique française 37. Brussels/Frankfurt am Main/New York: Peter Lang.
- Horsella, Maria, and Fresia Pérez. 1991. "Nominal Compounds in Chemical English Literature: Towards an Approach to Text Typology." *English for Specific Purposes* 10 (2): 125–138. [https://doi.org/10.1016/0889-4906\(91\)90005-H](https://doi.org/10.1016/0889-4906(91)90005-H)
- Huang, Chu-ren, Nicoletta Calzolari, Aldo Gangemi, Alessandro Lenci, Alessandro Oltramari, and Laurent Prévot (eds.) 2010. *Ontology and the Lexicon. A Natural Language Perspective*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511676536>
- Humbert-Droz, Julie. 2021. "Circulation des termes entre langues de spécialité et langue générale : proposition d'un cadre théorique et méthodologique d'analyse du phénomène de la déterminologisation." PhD diss., University of Geneva and University of Toulouse Jean-Jaurès.
- Humbert-Droz, Julie, Aurélie Picton, and Anne Condamines. 2019. "How to Build a Corpus for a Tool-based approach to Terminologisation in the Field of Particle Physics." *Research in Corpus Linguistics* (7): 1–17. <https://doi.org/10.32714/ricl.07.01>
- Humbley, John. 1989. "Terminologie et conscience linguistique." *La banque des mots* no spécial: 97–104.
- Humbley, John. 1997. "Is Terminology Specialized Lexicography? The Experience of French-speaking Countries." *Hermes, Journal of Linguistics* 18: 13–31.
- Humbley, John. 2000. "La terminologie." In *Histoire de la langue française 1945–2000*, edited by Gérard Antoine, and Bernard Cerquiglini, 315–338. Paris: CNRS Editions.
- Humbley, John. 2002. "Retour aux origines de la terminologie : l'acte de dénomination." *Langue française* 174 (2): 111–125.
- Humbley, John. 2003. "La néologie en terminologie." *Les Cahiers du Centre du Français Moderne* 2003 (1): 3–6.
- Humbley, John. 2004. "La réception de l'œuvre d'Eugen Wüster dans les pays de langue française." *Cahiers du C.I.E.L. 2004–2005*: 33–51. (<https://halshs.archives-ouvertes.fr/halshs-00276087>). Accessed 7 December 2021.
- Humbley, John. 2006. "Metaphor and Secondary Term Formation." In *La métaphore: du discours général aux discours spécialisés. Cahiers du C.I.E.L. 2000–2003*: 199–212.
- Humbley, John. 2007. "Vers une réception plurielle de la théorie terminologique de Wüster: une lecture commentée des avant-propos du manuel *Einführung in die allgemeine Terminologielehre*." *Langages* 168: 82–91. <https://doi.org/10.3917/lang.168.0082>
- Humbley, John. 2015. "Terminologie et traduction: relations avec la linguistique appliquée." In *Cultures de recherche en linguistique appliquée*, edited by Francis Carton, Jean Paul Narcy-Combes, Marie-Françoise Narcy-Combes, and Denyse Toffoli, 69–76. Série Langues et perspectives didactiques. Paris: Riveneuve Edition.
- Humbley, John. 2017. "La terminologie en tant que linguistique appliquée." Workshop in the *The Fourteenth International Conference on The History of the Language Sciences (ICHoLS XIV)*, Paris, August 28– September 1, 2017.
- Humbley, John. 2018a. "Term Formation and Neology." In *Language for Special Purposes*, edited by John Humbley, Gerhard Budin, and Christer Laurén, 437–452. Berlin/New York: Mouton de Gruyter. <https://doi.org/10.1515/9783110228014-022>
- Humbley, John. 2018b. *La néologie terminologique*. Limoges: Lambert-Lucas.
- Humbley, John, Gerhard Budin, and Christer Laurén (eds.). 2018. *Language for Special Purposes: An International Handbook*. Berlin/Boston: De Gruyter Mouton. <https://doi.org/10.1515/9783110228014>

- Humbley, John, and Aurélie Picton. 2017. "Introduction. Multiple Perspectives on Terminological Variation." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francœur, John Humbley, and Aurélie Picton, 1–7. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.00iint>
- Hunt, Elle. 2016. "Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter." *The Guardian*, March 24, 2016. (<https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism-from-twitter>). Accessed 7 December 2021.
- Huser, Vojtech. 2016. "The Growth and Evolution of Clinical Terminologies: Comparison of SNOMED, LOINC and RxNorm: Have They Come of Age?" Poster presented at the National Institute of Health Research Festival, Bethesda, MD, USA. (<https://researchfestival.nih.gov/2016/posters/growth-and-evolution-clinical-terminologies-comparison-snomed-loinc-and>). Accessed 7 December 2021.
- Huttin, Mathilde. 2015. "Loanword Adaptation: Phonological and Cognitive Issues." *Yearbook of the German Cognitive Linguistics Association* 3 (1), 89–104. <https://doi.org/10.1515/gcla-2015-0006>
- Hwang, Mi-Nyeong, Min-Hee Cho, Myung-Gwon Hwang, Mikyoung Lee, and Do-Heon Jeong. 2011. "Application of Trend Detection of Technical Terms to Technology Opportunity Discovery." In *U- and E-Service, Science and Technology, Proceedings of the International Conference UNESST 2011*, edited by Tai-hoon Kim, Hojjat Adeli, Jianhua Ma, Wai-chi Fang, Byeong-ho Kang, Byungjoo Park, Frode Eika Sandnes, and Kun Chang Lee, 258–262. Berlin: Springer. https://doi.org/10.1007/978-3-642-27210-3_33
- Hyland, Ken. 2002. "Genre: Language, Context, and Literacy." *Annual Review of Applied Linguistics* 22: 113–135. <https://doi.org/10.1017/S0267190502000065>
- Hyland, Ken, Meng Huat Chau, and Michael Handford. 2012. *Corpus Applications in Applied Linguistics*. London/New York: Continuum International Publishing Group.
- I**
- Ilinca, Elena-Cristina. 2013. "La terminologie : quelques repères dans la formation d'une discipline." *Studii și cercetări filologice. Seria Limbi Străine Aplicate* 12: 189–198.
- J**
- Jacquemin, Christian. 2001. *Spotting and Discovering Terms through Natural Language Processing*. Cambridge, MA: MIT Press.
- Jacques, Marie-Paule. 2000. "La réduction du syntagme terminologique au fil du discours." *Cahiers de grammaire* 25: 93–114.
- Jacques, Marie-Paule, and Nathalie Aussenac-Gilles. 2006. "Variabilité des performances des outils de TAL et genre textuel: Cas des patrons lexico-syntactiques." *Traitement automatique des langues* 47 (1): 11–42.
- Jacques, Marie-Paule, and Agnès Tutin. 2015. "Termes linguistiques et lexique transdisciplinaire des sciences humaines : observations en contexte." In *TOTH (Terminologie et Ontologies, Théories et applications)*. (<https://hal.archives-ouvertes.fr/hal-01903424>). Accessed 7 December 2021.
- Jakubíček, Milos, Adam Kilgarriř, Diana McCarthy, and Pavel Rychlý. 2010. "Fast Syntactic Searching in Very Large Corpora for Many Languages." In *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation (PACLIC) 24*, edited by Ryo Otoguro, Kiyoshi Ishikawa, Hiroshi Umemoto, and Yasunari Harada, 741–747. Tokyo: Waseda University.

- Janssen, Maarten. 2003. "Lexical Translation and Conceptual Hierarchies." In *Proceedings of the 5th International Symposium on Language, Logic and Computation*, Tbilisi, Georgia. (<http://maarten.janssenweb.net/Papers/tbilisi03-mjanssen.pdf>). Accessed 7 December 2021.
- Janssen, Maarten, and Marc Van Campenhoudt. 2005. "Terminologie traductive et représentation des connaissances : l'usage des relations hyponymiques." *Langages* 157: 63–79.
- Jared, Carlos, Alasdir Burman, and Nelson Léon. 1997. "Les amphisbènes, folklore et biologie." *La Recherche* 302: 38–40.
- Johns, Tim. 1991. "From Printout to Handout: Grammar and Vocabulary Teaching in the Context of Data-driven Learning." In *Classroom Concordancing*, edited by Tim Johns, and Philip King, 27–45. Vol. 4 of *English Language Research Journal*. Birmingham: Centre for English Language Studies, University of Birmingham.
- Johnson, Christopher. 1997. "Metaphor vs. Conflation in the Acquisition of Polysemy: The Case of SEE." In *Cultural, Psychological and Typological Issues in Cognitive Linguistics. Current Issues in Linguistic Theory* 152, edited by Masako Hiraga, Chris Sinha, and Sherman Wilcox, 155–170. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/cilt.152.12joh>
- Johnson, Mark. 2015. "Embodied Understanding." *Frontiers in Psychology* 6: 875. <https://doi.org/10.3389/fpsyg.2015.00875>
- Johnson, R. L., and Juan C. Sager. 1980. "Standardization of Terminology in a Model of Communication." *International Journal of the Sociology of Language* 23 (1980): 81–104. <https://doi.org/10.1515/ijsl.1980.23.81>
- Jones, Clive, John Lawton, and Moshe Shachak. 1994. "Organisms as Ecology Engineers." *Oikos* 69 (3): 373–386. <https://doi.org/10.2307/3545850>
- Josselin-Leray, Amélie, Cécile Fabre, Josette Rebeyrolle, Aurélie Picton, and Emmanuel Planas. 2014. "Good Contexts for Translators – A First Account of the CRISTAL Project." In *Proceedings of the XVI EURALEX International Congress: The User in Focus*, edited by Andrea Abel, Chiara Vettori, and Natascia Ralli, 631–645. Bolzano: Institute for Specialised Communication and Multilingualism.
- Justeson, John, and Slava M. Katz. 1995. "Technical Terminology: Some Linguistic Properties and an Algorithm for Identification in Text." *Natural Language Engineering* 1 (1): 9–27. <https://doi.org/10.1017/S1351324900000048>

K

- Kadima Batumona Adi, Gilbert. 2013. "Etude de la néologie terminologique en Bantu: cas de la cilubalisation des termes français des équipements et matériels de bureau." *Afrika Focus* 26(1): 108–128. <https://doi.org/10.1163/2031356X-02601009>
- Kageura, Kyo. 1996. "Methods of Automatic Term Recognition – A Review." *Terminology* 3(2): 259–289. <https://doi.org/10.1075/term.3.2.03kag>
- Kageura, Kyo. 1997a. "On Intra-Term Relations of Complex Terms in the Description of Term Formation Patterns." In *Mélanges de Linguistique offerts à Rostislav Kocourek*, edited by Lise Lapierre, Irène Oore, and Hans R. Runte, 105–111. Halifax: Presses d'ALFA.
- Kageura, Kyo. 1997b. "Multifaceted/Multidimensional Concept Systems." In *Handbook of Terminology Management: Basic Aspects of Terminology Management*, edited by Sue Ellen Wright, and Gerhard Budin, 119–132. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.htm1.15kag>
- Kageura, Kyo. 2002. *The Dynamics of Terminology*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/trlp.5>
- Kageura, Kyo. 2012. *The Quantitative Analysis of the Dynamics and Structure of Terminologies*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/trlp.15>

- Kageura, Kyo. 2015. "Terminology and Lexicography." In *Handbook of Terminology*, Vol. 1, edited by Hendrik J. Kockaert, and Frieda Steurs, 45–59. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/hot.1.04ter2>
- Kageura, Kyo. 2019. "What Does MT Research Say about Translation? An Empirical Sketch." In *Proceedings of the 2019 ERITS Conference*, 17–21. Seoul, Korea.
- Kageura, Kyo. 2020. "The Status and Role of Dictionaries in the Era of Unlimited Online Lexical Information Search and Free Online MT Use." *Journal of Korealex* 36: 7–33.
- Kageura, Kyo, and Long-Huei Chen. 2019. "Entropic Characterisation of Termino-Conceptual Structure: A Preliminary Study." In *Proceedings of the 12th International Conference on Terminology and Artificial Intelligence (TIA 2019)*, 55–68. Toulouse, France.
- Kageura, Kyo, Yasuko Enosawa, Miguel Santana-Rodriguez, and Sandra Yamilet-Santana. 1999. "A Comparative Observation of English, Spanish and Japanese Terminological Structure." In *Terminology and Knowledge Engineering 1999, Proceedings of the Fifth International Congress on Terminology and Knowledge Engineering*, edited by Peter Sandrini, 726–737. Vienna: TermNet.
- Kageura, Kyo, and Marie-Claude L'Homme. 2008. "Reflecting on Fifteen Years of Research and Development in Terminology." *Terminology* 14 (2): 153–157. <https://doi.org/10.1075/term.14.2.01edi>
- Kageura, Kyo, and Elizabeth Marshman. 2019. "Terminology Extraction and Terminology Management." In *Routledge Handbook of Translation and Technology*, edited by Minako O'Hagan, 61–77. London: Routledge. <https://doi.org/10.4324/9781315311258-4>
- Kamel, Mouna, Cassia Trojahn, Adel Ghamnia, Nathalie Aussenac-Gilles, and Cécile Fabre. 2017. "A Distant Learning Approach for Extracting Hypernym Relations from Wikipedia Disambiguation Pages." *Procedia Computer Science* 112: 1764–1773. <https://doi.org/10.1016/j.procs.2017.08.208>
- Katsurai, Marie, and Shunsuke Ono. 2019. "TrendNets: Mapping Emerging Research Trends from Dynamic Co-Word Networks via Sparse Representation." *Scientometrics* 121: 1583–1598. <https://doi.org/10.1007/s11192-019-03241-6>
- Katz, Slava M. 1987. "Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer." *IEEE Transactions on Acoustics, Speech and Signal Processing* 35: 400–401. <https://doi.org/10.1109/TASSP.1987.1165125>
- Kawaguchi, Yuji, Makoto Minegishi, and Wolfgang Viereck (eds). 2011. *Corpus-based Analysis and Diachronic Linguistics*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tufs.3>
- Kerremans, Koen. 2010. "A Comparative Study of Terminological Variation in Specialised Translation." In *Reconceptualizing LSP. Online Proceedings of the XVII European LSP Symposium 2009*, edited by Carmen Heine, and Jan Engberg. Aarhus: Aarhus School of Business, Aarhus University.
- Kerremans, Koen. 2017. "Towards a Resource of Semantically and Contextually Structured Term Variants and their Translations." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francœur, John Humbley, and Aurélie Picton, 83–108. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.04ker>
- Kerremans, Koen, Peter De Baer, and Rita Temmerman. 2010. "Competency-based Job Descriptions and Termontography. The Case of Terminological Variation." In *Terminology in Everyday Life*, edited by Marcel Thelen, and Frida Steurs, 181–194. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.13.13ker>

- Kerremans, Koen, and Rita Temmerman. 2004. Towards Multilingual, Termonological Support in Ontology Engineering. In *Proceedings of the Workshop on Terminology, Ontology and Knowledge representation*. Lyon, France. (https://www.academia.edu/1114290/Towards_multilingual_terminological_support_in_ontology_engineering). Accessed 7 December 2021.
- Kerremans, Koen, and Rita Temmerman. 2016. "Finding the (Un)expected? Quantitative and Qualitative Comparisons of Term Variants and their Translations in a Parallel corpus of EU texts." In *Corpus-based Approaches to Translation and Interpreting: from Theory to Applications*, edited by Gloria Corpas-Pastor, and Miriam Seghiri, 43–63. Frankfurt/Main: Peter Lang.
- Kerremans, Koen, Rita Temmerman, and Vanessa Andries. 2016. "Studying the Dynamics of Understanding and Legal Neologisms within a Linguistically Diverse Judicial Space: The Case of Motherhood in Belgium." *Procedia – Social and Behavioral Sciences* 231: 46–52. <https://doi.org/10.1016/j.sbspro.2016.09.070>
- Kerremans, Koen, Rita Temmerman, and Peter De Baer. 2008. "Construing Domain Knowledge via Terminological Understanding." *Linguistica Antverpiensia* 7: 177–191.
- Kiefer, Markus, and Friedemann Pulvermüller. 2012. "Conceptual Representations in Mind and Brain: Theoretical Developments, Current Evidence and Future Directions." *Cortex* 48: 805–825. <https://doi.org/10.1016/j.cortex.2011.04.006>
- Kiela, Douwe, and Stephen Clark. 2014. "A Systematic Study of Semantic Vector Space Model Parameters." In *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality (CVSC)* at the European Chapter of the Association for Computational Linguistics (EACL), edited by Alexandre Allauzen, Raffaella Bernardi, Edward Grefenstette, Hugo Larochelle, Christopher Manning, and Scott Wen-tau Yih, 21–30. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-1503>
- Kilgariff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, and Vít Suchomel. 2014. "The Sketch Engine: Ten Years On." *Lexicography* 1: 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kim, Minchai. 2017. "Variation terminologique en francophonie. Élaboration d'un modèle d'analyse des facteurs d'implantation terminologique." PhD diss., Université Paris IV.
- Kleiber, Georges. 1984. "Dénominations et relations dénominales." *Langages* 76: 77–94. <https://doi.org/10.3406/lgqe.1984.1496>
- Knapp, Karlfried. 2011. "Angewandte Linguistik in Deutschland: eine Disziplin?" *Histoire Epistémologie Langage* 33 (1): 117–128. <https://doi.org/10.3406/hel.2011.3210>
- Knudsen, Susanne. 2003. "Scientific Metaphors Going Public." *Journal of Pragmatics* 35: 1247–1263. [https://doi.org/10.1016/S0378-2166\(02\)00187-X](https://doi.org/10.1016/S0378-2166(02)00187-X)
- Knudsen, Susanne. 2005. "Communicating Novel and Conventional Scientific Metaphors: A Study of the Development of the Metaphor of Genetic Code." *Public Understanding of Science* 14 (4): 373–392. <https://doi.org/10.1177/0963662505056613>
- Koby, Geoffrey S. 2015. "The ATA Flowchart and Framework as a Differentiated Error-Marking Scale in Translation Teaching." In *Handbook of Research on Teaching Methods in Language Translation and Interpretation*, edited by Ying Cui, and Zhao Wei, 220–253. Hershey, PA: IGI Global. <https://doi.org/10.4018/978-1-4666-6615-3.ch013>
- Kocourek, Rostislav. 1968. "Synonymy and Semantic Structures of Terminology." *Travaux Linguistiques de Prague* 3: 131–141.
- Kocourek, Rostislav. 1991. *La langue française de la technique et de la science. Vers une linguistique de la langue savante*. Wiesbaden: Oscar Brandstetter.
- Koller, Veronika. 2004. *Metaphor and Gender in Business Media Discourse: A Critical Cognitive Study*. Basingstoke/New York: Palgrave Macmillan. <https://doi.org/10.1057/9780230511286>

- Koller, Veronika. 2006. "Of Critical Importance: Using Corpora to Study Metaphor in Business Media Discourse." In *Corpus-Based Approaches to Metaphor and Metonymy. Trends in Linguistics. Studies and Monographs*, edited by Anatol Stefanowitsch, and Stefan T. Gries, 229–257. Berlin/New York: Mouton de Gruyter.
- Kövecses, Zoltan. 2002. *Metaphor: A Practical Introduction*. Oxford: Oxford University Press.
- Kövecses, Zoltan. 2005. *Metaphor in Culture. Universality and Variation*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511614408>
- Kövecses, Zoltan. 2015. *Where Metaphors Come from. Reconsidering Context in Metaphor*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780190224868.001.0001>
- Kövecses, Zoltan, Laura Ambrus, Daniel Hegedű, Ren Imai, and Anna Sobczak. 2019. "The Lexical vs. Corpus-Based Method in the Study of Metaphors." In *Metaphor and Metonymy in the Digital Age*, edited by Marianna Bolognesi, Mari Brdar, and Kristina Despot, 149–173. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/milcc.8.07kov>
- Kozerenko, Elena B., Nikolai V. Somin, Yulia I. Sinyaghina, and Svetlana A. Pozdeeva. 2019. "Problem Domain Ontology Mining Based on Distributional Semantics." In *Proceedings of the 6th International Conference on Computational Science and Computational Intelligence (CSCI)*, edited by Hamid R. Arabnia, Leonidas Deligiannidis Tran, 439–444. Washington, DC/Los Alamitos, CA: Conference Publishing Services. <https://doi.org/10.1109/CSCI49370.2019.00086>
- Kozinets, Robert. 2015. *Netnography: Redefined*. 2nd ed. London: Sage Publications.
- Krishna, Aradhna, and Norbert Schwarz. 2014. "Sensory Marketing, Embodiment, and Grounded Cognition: A Review and Introduction." *Journal of Consumer Psychology* 24 (2): 159–68. <https://doi.org/10.1016/j.jcps.2013.12.006>
- Kübler, Natalie. 2001. "Corpora in Terminology and Translation teaching: methodological approach." In *Proceedings of the ICAME 01 Conference (May 2001), Extending the Scope of Corpus-based Research: New Applications, New Challenges*, edited by Sylviane Granger, and Stephanie Petch-Tyson, 53–55. Amsterdam: Rodopi.
- Kübler, Natalie. (2003) 2014. "Corpora and LSP Translation." In *Corpora in Translator Education*, edited by Federico Zanettin, Silvia Bernardini, and Dominic Stewart, 25–42. Abingdon-on-Thames, UK: Routledge.
- Kübler, Natalie, Alexandra Mestivier-Volanschi, and Mojca Pecman. 2022, forthcoming. "Using Comparable Corpora for Translating and Post-editing Complex Noun Groups in Specialised Texts (from English to French)." In *Extending the Scope of Corpus-based Translation Studies*, edited by Sylviane Granger, and Marie-Aude Lefer, 237–266. New York: Bloomsbury.
- Kuhn, Thomas. 1979. "Metaphor in Science." In *Metaphor and Thought*, edited by Andrew Ortony, 409–419. Cambridge: Cambridge University Press.
- Kuteeva, Maria, and Raffaella Negretti. 2016. "Graduate Students' Genre Knowledge and Perceived Disciplinary Practices: Creating a Research Space across Disciplines." *English for Specific Purposes* 41: 36–49. <https://doi.org/10.1016/j.esp.2015.08.004>
- Kytö, Merja, Juhani Rudanko, and Erik Smittberg. 2000. "Building a Bridge between the Present and the Past: A Corpus of 19th-century English." *ICAME Journal* 25: 82–98.

L

- Lafourcade, Mathieu, and Lionel Ramadier. 2016. "Semantic Relation Extraction with Semantic Patterns Experiment on Radiology Reports." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2016)*, 4578–4582. Portoroz, Slovenia: European Language Resources Association.

- Lakatos, Imre. 1978. *The Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511621123>
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things, What Categories Reveal about the Mind*. Chicago: Chicago University Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Lakoff, George. 1993. "The Contemporary Theory of Metaphor". In *Metaphor and Thought*, edited by Andrew Ortony, 202–251. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139173865.013>
- Lakoff, George, and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Lakoff, George, and Mark Johnson. 2003 [1980]. *Metaphors We Live By*. Chicago: Chicago University Press.
- Lakoff, George, and Mark Turner. 1989. *More than Cool Reason: A Field Guide to Poetic Metaphor*. Chicago: University of Chicago. <https://doi.org/10.7208/chicago/9780226470986.001.0001>
- Lambon Ralph, Matthew A., Elisabeth Jefferies, Karalyn Patterson, and Timothy T. Rogers. 2017. "The Neural and Computational Bases of Semantic Cognition." *Nature Reviews Neuroscience* 18 (1): 42–55. <https://doi.org/10.1038/nrn.2016.150>
- Lambon Ralph, Matthew A., Karen Sage, Roy W. Jones, and Emily J. Mayberry. 2010. "Coherent Concepts are Computed in the Anterior Temporal Lobes." *PNAS* 107 (6): 2717–2722. <https://doi.org/10.1073/pnas.0907307107>
- Landauer, Thomas K., and Susan T. Dumais. 1997. "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Introduction and Representation of Knowledge." *Psychological Review* 104 (2): 211–240. <https://doi.org/10.1037/0033-295X.104.2.211>
- Lang, Friedrich H. 1998. "Eugen Wüster zum 100. Geburtstag. Ein Elektrotechniker als Terminologe." *Elektrotechnik und Informationstechnik* 115 (11): 625–628. <https://doi.org/10.1007/BF03161106>
- Langacker, Ronald. 1987. *Foundations of Cognitive Grammar, Volume I, Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Lapesa, Gabriella, and Stefan Evert. 2014. "A Large-scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection." *Transactions of the Association for Computational Linguistics* 2: 531–545. https://doi.org/10.1162/tacl_a_00201
- Laurén, Christer, Johan Myking, and Heribert Picht. 1998. *Terminologie unter der Lupe*. Vienna: TermNet.
- Laurén, Christer, and Heribert Picht. 1993. "Vergleich der terminologischen Schulen." In *Ausgewählte Texte zur Terminologie*, edited by Christer Laurén, and Heribert Picht, 493–539. Vienna: TermNet, Infoterm.
- Lavagnino, Elisa. 2012. "Les Langues de Spécialité et la variation discursive: une étude contrastive de la réduction des termes complexes." *Synergies Espagne* 5: 47–64.
- Ledouble, Hélène. 2020. "Term Circulation and Conceptual Instability in the Mediation of Science: Binary Framing of the Notions of Biological versus Chemical Pesticides." *Discourse & Communication* 14 (5): 466–488. <https://doi.org/10.1177/1750481320917575>
- Lee, Jinhyuk, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. "BioBERT: A Pre-trained Biomedical Language Representation Model for Biomedical Text Mining." *Bioinformatics*. arXiv:1901.08746 [cs.CL]. <https://doi.org/10.1093/bioinformatics/btz682>
- Lefever, Els, Marjan Van de Kauter, and Véronique Hoste. 2014. "HypoTerm: Detection of Hyponym Relations between Domain-specific Terms in Dutch and English." *Terminology* 20 (2): 250–278. <https://doi.org/10.1075/term.20.2.061ef>

- Lemaire, Benoît, and Guy Denhière. 2006. "Effects of High-Order Co-occurrences on Word Semantic Similarity." *Current Psychology Letters* 18(1). <https://doi.org/10.4000/cpl.471>
- Lenci, Alessandro. 2008. "Distributional Approaches in Linguistic and Cognitive Research." *Italian Journal of Linguistics* 20: 1–31.
- Lenci, Alessandro. 2018. "Distributional Models of Word Meaning." *Annual Review of Linguistics* 4: 151–171. <https://doi.org/10.1146/annurev-linguistics-030514-125254>
- León-Araúz, Pilar. 2009. "Representación multidimensional del conocimiento especializado: el uso de marcos desde la macroestructura hasta la microestructura." PhD diss. University of Granada.
- León-Araúz, Pilar. 2014. "Semantic Relations and Local Grammars for the Environment." In *Formalising Natural Languages with NooJ 2013*, edited by Svetla Koeva, Slim Mesfar, and Max Silberstein, 87–102. Newcastle-on-Tyne: Cambridge Scholars.
- León-Araúz, Pilar. 2017. "Term and Concept Variation in Specialized Knowledge Dynamics." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francoeur, John Humbley, and Aurélie Picton, 213–258. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.09leo>
- León-Araúz, Pilar, Melania Cabezas-García, Arianne Reimerink. 2020. "Representing Multiword Term Variation in a Terminological Knowledge Base: a Corpus-Based Study." In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, 2351–2360. Luxemburg: European Language Resources Association.
- León-Araúz, Pilar, and Pamela Faber. 2010. "Natural and Contextual Constraints for Domain-specific Relations." In *Proceedings of the Workshop Semantic Relations. Theory and Applications*, edited by Verginica Barbu Mititelu, Viktor Pekar, and Eduard Barbu, 12–17. Valletta, Malta. (<http://www.lrec-conf.org/proceedings/lrec2010/workshops/W9.pdf>). Accessed 7 December 2021
- León Araúz, Pilar, and Pamela Faber. 2014. "Context and Terminology in the Multilingual Semantic Web." In *Towards the Multilingual Semantic Web*, edited by Philipp Cimiano, and Paul Buitelaar, 31–47. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-662-43585-4_3
- León Araúz, Pilar, and Arianne Reimerink. 2010. "Knowledge Extraction and Multidimensionality in the Environmental Domain." In *Proceedings of the 9th International Conference on Terminology and Knowledge Engineering (TKE)*. CD-ROM. Dublin: Dublin City University.
- León-Araúz, Pilar, and Arianne Reimerink. 2014. "From Term Dynamics to Concept Dynamics: Term Variation and Multidimensionality in the Psychiatric Domain." In *Proceedings of the XVI EURALEX International Congress*, edited by Andrea Abel, Chiara Vettori, and Nataschia Ralli, 657–667. Bolzano, Italy.
- León-Araúz, Pilar, and Arianne Reimerink. 2018. "Improved Knowledge-Rich Context Extraction for Terminography." In *Technological Innovation for Specialized Linguistic Domains*, edited by Timothy Read, Salvador Montaner, and Beatriz Sedano, 69–84. Saarbrücken, Germany: Éditions Universitaires Européennes.
- León-Araúz, Pilar, and Arianne Reimerink. 2019. "High-Density Knowledge-Rich Contexts." *Argentinian Journal of Applied Linguistics* 7 (1): 109–130.
- León-Araúz, Pilar, Arianne Reimerink, and Pamela Faber. 2013. "Multidimensional and Multimodal Information in EcoLexicon." In *Computational Linguistics*, edited by Adam Przepiórkowski, Maciej Piasecki, Krzysztof Jassem, and Piotr Fuglewicz, 143–161. Berlin: Springer. https://doi.org/10.1007/978-3-642-34399-5_8

- León-Araúz, Pilar, Arianne Reimerink, and Pamela Faber. 2019. "Assessing EcoLexiCAT: Terminology Enhancement and Post-editing." In *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2019 Conference*, edited by Iztok Kosem, Tanara Zingano-Kuhn, Margarita Correia, José Pedro Ferreira, Maarten Jansen, Isabel Pereira, Jelena Kallas, Milos Jakubiček, Simon Krek, and Carole Tiberius, 130–160. Brno: Lexical Computing CZ, s.r.o.
- León-Araúz, Pilar, Arianne Reimerink, and Alejandro García-Aragón. 2013. "Dynamism and Context in Specialized Knowledge." *Terminology* 19 (1): 31–61.
<https://doi.org/10.1075/term.19.1.02leo>
- León-Araúz, Pilar, and Antonio San Martín. 2018. "The EcoLexicon Semantic Sketch Grammar: From Knowledge Patterns to Word Sketches." In *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*, edited by Jaka Čibej, Vojko Gorjanc, Iztok Kosem, and Simon Krek, 94–99. Ljubljana: Ljubljana University Press.
- León-Araúz, Pilar, Antonio San Martín, and Pamela Faber. 2016. "Pattern-Based Word Sketches for the Extraction of Semantic Relations." In *Proceedings of the 5th International Workshop on Computational Terminology (Computerm2016)*, edited by Patrick Drouin, Natalia Grabar, Thierry Hamon, Kyo Kageura, and Koichi Takeuchi, 73–82. Osaka: Japan.
- León-Araúz, Pilar, Antonio San Martín, and Arianne Reimerink. 2018. "The EcoLexicon English Corpus as an Open Corpus in Sketch Engine." In *Proceedings of the XVIII EURALEX International Congress. Lexicography in Global Contexts*, edited by Jaka Čibej, Vojko Gorjanc, Iztok Kosem, and Simon Krek, 893–901. Ljubljana: Ljubljana University Press.
- Lerat, Pierre. 1984. "Lexicologie des institutions." In *Sociolinguistique des langues romanes. Actes Du XVIIe Congrès international de linguistique et de philologie romanes*, 251–259. Aix-en-Provence: Université de Provence.
- Lerat, Pierre. 1988. "Terminologie et sémantique descriptive." *La Banque des mots*, numéro spécial 1 CTN: 11–29.
- Lerat, Pierre. 1989. "Les fondements théoriques de la terminologie." *La Banque des mots*, numéro spécial 2 CTN: 51–61.
- Lerat, Pierre. 2000. "Le sens terminologique." In *Louis Guespin, terminologue. Actes de la journée en hommage à Louis Guespin, terminologue*, edited by Valérie Delavigne, 17–25. Rouen: Presses de l'Université de Rouen.
- Lerat, Pierre. 2002. "Qu'est-ce que le verbe spécialisé? Le cas du droit." *Cahiers de lexicologie* 80 (1): 201–211.
- Leroy, Nathalie. 2005. "L'universalisation des droits de l'homme. Approche sociolinguistique des débats des Nations-Unies." PhD diss., Université de Rouen.
- Lervad, Susanne. 2009. "Recherche en terminologie et applications pratiques: quelques axes de collaboration avec des partenaires industriels danois." *Revue française de linguistique appliquée* XIV: 73–80. <https://doi.org/10.3917/rfla.142.0073>
- Lethuiller, Jacques. 1989. "La synonymie en langue de spécialité." *Meta* 34 (3): 443–449.
<https://doi.org/10.7202/003498ar>
- Levi, Judith. 1978. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
- Levin, Beth. 2013. "Argument Structure." In *Oxford Bibliographies in Linguistics*, edited by Mark Aronoff. Oxford: Oxford University Press. <https://doi.org/10.1093/obo/9780199772810-0099>
- Levy, Omer, and Yoav Goldberg. 2014. "Dependency-Based Word Embeddings." In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL) (Vol. 2: Short Papers)*, edited by Kristina Toutanova, and Hua Wu, 302–308. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/P14-2050>

- Levy, Omer, Yoav Goldberg, and Ido Dagan. 2015. "Improving Distributional Similarity with Lessons Learned from Word Embeddings." *Transactions of the Association for Computational Linguistics* 3: 211–225. https://doi.org/10.1162/tacl_a_00134
- L'Homme, Marie-Claude. 1998. "Le statut du verbe en langue de spécialité et sa description lexicographique." *Cahiers de lexicologie* 73 (2): 61–84.
- L'Homme, Marie-Claude. 2003. "Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivatives: A Model for Specialized Lexicography." *International Journal of Lexicography* 16 (4): 403–422. <https://doi.org/10.1093/ijl/16.4.403>
- L'Homme, Marie-Claude. 2004. *La terminologie: principes et techniques*. Montreal: Presses de l'Université de Montréal. <https://doi.org/10.4000/books.pum.10693>
- L'Homme, Marie-Claude. 2006a. "The Processing of Terms in Dictionaries: New Models and Techniques: A State of the Art." *Terminology* 12 (2): 181–188. <https://doi.org/10.1075/term.12.2.02hom>
- L'Homme, Marie-Claude. 2006b. A Look at some Canadian Contributions to Terminology. In *Modern Approaches to Terminological Theories and Applications*, edited by Heribert Picht, 55–77. Switzerland: Peter Lang.
- L'Homme, Marie-Claude. 2012. "Adding Syntactico-Semantic Information to Specialized Dictionaries: An Application of the FrameNet Methodology." *Lexicographica* 28 (1): 233–252. <https://doi.org/10.1515/lexi.2012-0012>
- L'Homme, Marie-Claude. 2018. "Maintaining the Balance between Knowledge and the Lexicon in Terminology: A Methodology Based on Frame Semantics." *Lexicography* 4 (1): 3–21. <https://doi.org/10.1007/s40607-018-0034-1>
- L'Homme, Marie-Claude. 2020a. *Lexical Semantics for Terminology: An Introduction*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.20>
- L'Homme, Marie-Claude. 2020b. "Revisiting Polysemy in Terminology." In *Proceedings of the XIX EURALEX*, edited by Zoe Gavriilidou, Maria Mitsiaki, and Asimakis Fliatouras, 415–424. Greece: Democritus University of Thrace.
- L'Homme, Marie-Claude. 2020c. "Being a Privileged Witness of Twenty Years of Research in Terminology. An editorial statement." *Terminology* 26 (1): 1–6. <https://doi.org/10.1075/term.00039.edi>
- L'Homme, Marie Claude, and Gabriel Bernier-Colborne. 2012. "Terms as Labels for Concepts, Terms as Lexical Units: A Comparative Analysis in Ontologies and Specialized Dictionaries." *Applied Ontology* 7(4): 387–400. <https://doi.org/10.3233/AO-2012-0116>
- L'Homme, Marie-Claude, and Elizabeth Marshman. 2006. "Terminological Relationships and Corpus-based Methods for Discovering Them: An Assessment for Terminographers." In *Lexicography, Terminology, and Translation. Text-based Studies in Honour of Ingrid Meyer*, edited by Lynne Bowker, 67–80. Ottawa: University of Ottawa Press. <https://doi.org/10.2307/j.ctt1ckpgs3.8>
- L'Homme, Marie-Claude, and Alain Polguère. 2008. "Mettre en bons termes les dictionnaires spécialisés et les dictionnaires de langue générale." In *Lexicographie et terminologie: histoire de mots. Hommage à Henri Béjoint*, edited by François Maniez, and Pascaline Dury, 191–206. Lyon: Presses de l'Université de Lyon.
- L'Homme, Marie-Claude, Benoit Robichaud, and Carlos Subirats. 2014. "Discovering Frames in Specialized Domains." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, edited by Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, 1364–1371. Paris: European Language Resources Association.

- Licata, Laurent, and Aurélie Mercy. 2015. "Collective Memory (Social Psychology of)." In *The International Encyclopedia of the Social and Behavioral Sciences*, edited by James Wright, 194–199. Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.24046-4>
- Liénard, Fabien, and Sami Zlitni (eds). 2015. *La communication électronique. Enjeux, stratégies, opportunités*. Limoges: Lambert Lucas.
- Lindblom, Jessica. 2015. "Meaning-making as a socially distributed and embodied practice." In *Aesthetics and the Embodied Mind: Beyond Art Theory and the Cartesian Mind-Body Dichotomy*, edited by Alfonsina Scarinzi, 3–19. Netherlands: Springer. https://doi.org/10.1007/978-94-017-9379-7_1
- Linn, Andrew. 2011. "Impact: Linguistics in the Real World." *Histoire Epistémologie Langage* 33 (1): 15–27. <https://doi.org/10.3406/hel.2011.3203>
- Lopez, Stéphanie. 2013. *Norme(s) et usage(s) langagiers: le cas des communications pilote-contrôleur en anglais*. Toulouse: Université Toulouse 2.
- López Arroyo, Belén, and Leticia Moreno Pérez. 2019. "Lexical Chunks in English and Spanish Sales Contracts." *Terminology* 25 (1): 32–59. <https://doi.org/10.1075/term.00027.lop>
- López-Rodríguez, Clara Inés, María Isabel Tercedor Sánchez, and Pamela Faber. 2006. "Gestión terminológica basada en el conocimiento y generación de recursos de información sobre el cáncer: el proyecto OncoTerm." *RevistaeSalud.com* 2 (8). (https://www.academia.edu/732009/Gestion_terminologica_basada_en_el_conocimiento_y_generacion_de_recursos_de_informacion_sobre_el_cancer_el_proyecto_Oncoterm). Accessed on 7 December 2021.
- López Rodríguez, Clara Inés, Miriam Buendía Castro, and Alejandro García Aragón. 2012. "User Needs to the Test: Evaluating a Terminological Knowledge Base on the Environment by Trainee Translators." *JoSTrans: The Journal of Specialised Translation* 18: 57–76.
- Lopopolo, Alessandro, and Emiel van Miltenburg. 2015. "Sound-based Distributional Models." In *Proceedings of the 11th International Conference on Computational Semantics*, edited by Matthew Purver, Mehrnoosh Sadrzadeh, and Matthew Stone, 70–75. Stroudsburg, PA: Association for Computational Linguistics.
- Lorente, Mercè. 2007. "Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació." In *Estudis de lingüístics i de lingüística aplicada en honor de M. Teresa Cabré Castellví*, edited by Mercè Lorente, Rosa Estopà, Judit Freixa, Jaume Martí, and Carles Tebé, 365–380. Volum II: De deixebles. Barcelona: Institut Universitari de Lingüística Aplicada de la Universitat Pompeu Fabra.

M

- Madsen, Bodil N., and Hanne E. Thomsen. 2009. "Terminological Concept Modelling and Conceptual Data Modelling." *International Journal of Metadata, Semantics and Ontologies (IJMSO)* 4 (4): 239–249. <https://doi.org/10.1504/IJMSO.2009.029228>
- Maedche, Alexander, and Steffen Staab. 2001. "Ontology Learning for the Semantic Web." *IEEE Intelligent Systems* 16 (2): 72–79. <https://doi.org/10.1109/5254.920602>
- Maguire, Phil, Edward J. Wisniewski, and Gert Storms. 2010. "A Corpus Study of Semantic Patterns in Compounding." *Corpus Linguistics and Linguistic Theory* 6 (1): 49–73. <https://doi.org/10.1515/cllt.2010.003>
- Mahesh, Kavi, and Sergei Nirenburg. 1995. "A Situated Ontology for Practical NLP." In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, International Joint Conference on Artificial Intelligence (IJCAI-1995)*, 1–10. Montreal, Canada.

- Maia, Belinda. (2003) 2014. "Training Translators in Terminology and Information Retrieval Using Comparable and Parallel Corpora" *Corpora in Translator Education*, edited by Federico Zanettin, Silvia Bernardini, and Dominic Stewart, 47–58. Abingdon-on-Thames, UK: Routledge.
- Maillot, Jean. 1969. *La traduction scientifique et technique*. Paris: Eyrolles.
- Mair, Christian. 1997. "Parallel Corpora: A Real-Time Approach to the Study of Language Change in Progress." In *Corpus-based Studies in English*, edited by Magnus Ljung, 195–209. Amsterdam: Rodopi.
- Malaisé, Véronique, Pierre Zweigenbaum, and Bruno Bachimont. 2005. "Mining Defining Contexts to Help Structuring Differential Ontologies." *Terminology* 11 (1): 21–53. <https://doi.org/10.1075/term.11.1.03mal>
- Mandin, Jérémy. 2014. *An Overview of Integration Policies in Belgium*. INTERACT Research Report, no. 2014/20. Florence: European University Institute and Robert Schuman Centre for Advanced Studies.
- Maniez, François. 2015. "The Use of Denominal Adjectives in Terminology." In *Terminología y comunicación científica y social*, edited by Mercedes Roldan Vendrell, 43–66. Granada: Comares.
- Manning, Christopher, and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.
- Marabini San Martín, Blanca. 2019. "Glottopolitics: The Power of Language." (https://www.ieee.es/Galerias/fichero/docs_opinion/2019/DIEEEO23_2019BLAMAR-gloto_ENG.pdf). Accessed 7 December 2021.
- Marcellesi, Jean-Baptiste, and Bernard Gardin. 1974. *Introduction à la sociolinguistique: la linguistique sociale*. Paris: Larousse.
- Marín, María José, and Camino Rea. 2014. "Researching Legal Terminology: A Corpus-based Proposal for the Analysis of Sub-technical Legal terms. Intersections – l'anglais de spécialité, creuset multidomaine." *ASP la revue du GERAS* 66: 1–22.
- Maroto, Nava Garcia, and Amparo Alcina. 2009. "Formal Description of Conceptual Relationships with a View to Implementing Them in the Ontology Editor Protégé". In *Terminology* 15 (2), 232–257. <https://doi.org/10.1075/term.15.2.04mar>
- Marshman, Elizabeth. 2002. "The Cause-effect Relation in Biopharmaceutical Texts: English Knowledge Patterns." In *Proceedings of Terminology and Knowledge Engineering* (TKE 2002), edited by Alain K. Melby, 89–94. Le Chesnay: Inria.
- Marshman, Elizabeth. 2007a. "Lexical Knowledge Patterns for Semi-automatic Extraction of Cause-effect and Association Relations from Medical Texts: A Comparative Study of English and French." PhD diss., Université de Montréal.
- Marshman, Elizabeth. 2007b. "Towards Strategies for Processing Relationships between Multiple Relation Participants in Knowledge Patterns: An Analysis in English and French." *Terminology* 13 (1): 1–34. <https://doi.org/10.1075/term.13.1.02mar>
- Marshman, Elizabeth. 2008. "Expressions of Uncertainty in Candidate Knowledge-rich Contexts: A Comparison in English and French Specialized Texts." *Terminology* 14 (1): 124–151. <https://doi.org/10.1075/term.14.1.07mar>
- Marshman, Elizabeth, Julie Gariépy, and Charissa Harms. 2012. "Helping Language Professionals Relate to Terms: Terminological Relations and Termbases." *The Journal of Specialised Translation* 18: 30–56.
- Marshman, Elizabeth, Marie-Claude L'Homme, and Victoria Surtees. 2008a. "Portability of Cause-effect Relation Markers across Specialised Domains and Text Genres: A Comparative Evaluation." *Corpora* 3 (2): 141–172. <https://doi.org/10.3366/E1749503208000130>

- Marshman, Elizabeth, Marie-Claude L'Homme, and Victoria Surtees. 2008b. "Verbal Markers of Cause-effect. Relations across Corpora." In *Managing Ontologies and Lexical Resources. Proceedings of the 8th International Conference on Terminology and Knowledge Engineering, TKE 2008*, edited by Bodil Nistrup Madsen, and Hanne Erdman Thomsen, 159–173. Copenhagen: Copenhagen Business School.
- Marshman, Elizabeth, Tricia Morgan, and Ingrid Meyer. 2002. "French Patterns for Expressing Concept Relations." *Terminology* 8 (1): 1–29. <https://doi.org/10.1075/term.8.1.02mar>
- Martin, André. 1994. "Théorie de la diffusion des innovations et implantation terminologique." *Terminologies Nouvelles* 11: 33–39.
- Martin, James Robert. 2002. "Meaning beyond the Clause: SFL Perspective." *Annual Review of Applied Linguistics* 22: 52–74. <https://doi.org/10.1017/S026719050200003X>
- Martin, James Robert, and David Rose. 2008. *Genre Relations: Mapping Culture*. London: Equinox.
- Martin, Louis, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamel Seddah, and Benoît Sagot. 2020. "CamemBERT: A Tasty French Language Model." In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*, edited by Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel Tetreault, 7203–7219. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.645>
- Martin, Willy, and Bernard P. F. Al. 1988. "User-Oriented in Dictionaries: 9 Propositions." In *Proceedings of the 3rd International Congress of the European Association for Lexicography*, 393–399. Budapest, Hungary.
- Martín-Chozas, Patricia, Sina Ahmadi, and Elena Montiel-Ponsoda. 2020. "Defying Wikidata: Validation of Terminological Relations in the Web of Data." In *Proceedings of the Language Resources and Evaluation Conference 2020*, 5654–5659. Marseille, France.
- Martín Mingorance, Leocadio. 1989. "Functional Grammar and Lexematics." In *Meaning and Lexicography*, edited by Jerzy Tomaszcyk, and Barbara Lewandowska, 227–253. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/jlsee.28.zomin>
- Matlock, Teenie. 2004. "Fictive Motion as Cognitive Simulation." *Memory and Cognition*, 32: 1389–1400. <https://doi.org/10.3758/BF03206329>
- Mayaffre, Damon. 2005. "Rôle et place du corpus en linguistique. Réflexions introductives." In *Actes Du Colloque JETOU'2005*, edited by Pascale Vergely, 5–17. Toulouse: Université de Toulouse-Le Mirail.
- McCrae, John, Guadalupe Aguado-de-Cea, Paul Buitelaar, Philipp Cimiano, Thierry Declerck, Asunción Gómez-Pérez, Jorge Gracia, Laura Hollink, Elena Montiel-Ponsoda, Dennis Spohr, and Tobias Wunner. 2012. "Interchanging Lexical Resources on the Semantic Web." In *Language Resources and Evaluation* 46 (4): 701–719. <https://doi.org/10.1007/s10579-012-9182-3>
- McCrae, John P., and Philipp Cimiano. 2018. *Linguistic Linked Open Data Cloud*. (<https://linguistic-lod.org/lod-cloud>). Accessed 7 December 2021.
- McGuinness, Deborah L. 2003. "Ontologies Come of Age." In *Spinning the Semantic Web: Bringing the World Wide Web to its Full Potential*, edited by Dieter Fensel, James A. Hendler, Henry Lieberman, and Wolfgang Wahlster, 171–194. Cambridge, MA: MIT Press.
- Medhat-Lecocq, Héba. 2014. "Objet, concept, terme, discours... Repenser le domaine dans le cadre de la terminologie comparée." In *TOTh [Terminologie et ontologie] 2014*. Chambéry: Université de Savoie (<https://hal.archives-ouvertes.fr/hal-01493382/document>). Accessed 7 December 2021.

- Medhat-Lecocq, Héba. 2019. "Le langage du droit : quels défis pour le traducteur juridique." *Langues, cultures et sociétés* 5 (1): 80–92.
- Mejri, Salah, Geneviève Petiot, and Gérard Petit. 2005. "Introduction." *Linx. Revue des linguistes de l'Université Paris X Nanterre* 52: 9–16. <https://doi.org/10.4000/linx.167>
- Meščuk, Igor, André Clas, and Alain Polguère. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot/Aupelf-UREF.
- Mensah, Barnabé. 2008. "La modernization du Gàngbe." PhD diss., University of Paris III.
- Messaoudi, Leila. 2002. "Opacité et transparence dans les technolectes bilingues (français-arabe)." *Meta* 45 (3): 424–436. <https://doi.org/10.7202/003641ar>
- Messaoudi, Leila. 2006. "Les rôles de la situation et du contexte dans les technolectes bilingues français-arabes." In *Mots, termes et contextes. Actes des septièmes journées scientifiques du réseau des chercheurs Lexicologie Terminologie Traduction*, edited by Daniel Blampain, Philippe Thoiron, Marc Van Campenhoudt, 687–697. Paris: Editions des archives contemporaines.
- Messineo, Cristina. 2002. "Variación conceptual y formal del término educación bilingüe intercultural (EBI) en distintos tipos de discursos." *Terminology* 8 (1): 113–19. <https://doi.org/10.1075/term.8.1.06mes>
- Meyer, Ingrid. 2001. "Extracting Knowledge-Rich Contexts for Terminography: A Conceptual and Methodological Framework." In *Recent Advances in Computational Terminology*, edited by Didier Bourigault, Christian Jacquemin, and Marie-Claude L'Homme, 279–302. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/nlp.2.15mey>
- Meyer, Ingrid, Lynne Bowker, and Karen Eck. 1992. "Cogniterm: An Experiment in Building a Terminological Knowledge Base." In *Proceedings of the 5th Euralex International Congress on Lexicography*, edited by Hannu Tommola, Krista Varantola, Tarja Salmi-Tolonen, and Jurgen Schopp, 159–172. Tampere: University of Tampere.
- Meyer, Ingrid, Karen Eck, and Douglas Skuce. 1997. "Systematic Concept Analysis within a Knowledge-based Approach to Terminology." In *Handbook of Terminology Management. Volume 1: Basic Aspects of Terminology Management*, edited by Sue Ellen Wright, and Gerhard Budin. 98–118. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.htm1.14mey>
- Meyer, Ingrid, and Kristen Mackintosh. 1996. "The Corpus from a Terminographer's Viewpoint." *International Journal of Corpus Linguistics* 1 (2): 257–285. <https://doi.org/10.1075/ijcl.1.2.05mey>
- Meyer, Ingrid, and Kristen Mackintosh. 2000. "When Terms Move in our Everyday Lives: An Overview of De-terminologization." *Terminology* 6 (1): 111–138. <https://doi.org/10.1075/term.6.1.07mey>
- Meyer, Ingrid, Kristen Mackintosh, Caroline Barrière, and Tricia Morgan. 1999. "Conceptual Sampling for Terminographical Corpus Analysis." In *Proceedings of Terminology and Knowledge Engineering (TKE '99)*, edited by Peter Sandrini, 256–267. Würzburg: Ergon-Verlag.
- Meyer, Ingrid, David Miller, and Diane Michaud. 1991. "Terminologie et analyse notionnelle assisté par ordinateur." In *Actes du colloque international sur les industries de la langue*, 187–199. Québec: Office québécois de la langue française.
- Meyer, Ingrid, and Line Paradis. 1991. "Applying Knowledge-Engineering Technology to Terminology: A Pilot Project." *Terminology Update* 24 (2): 3–8.
- Meyer, Ingrid, and Douglas Skuce. 1990. "Computer-Assisted Concept Analysis for Terminology: A Framework for Technological and Methodological Research." In *Fourth International Congress of the European Association for Lexicography*, Malaga, Spain.

- Meyer, Ingrid, Douglas Skuce, Lynne Bowker, and Karen Eck. 1992. "Towards a New Generation of Terminological Resources: An Experiment in Building a Terminological Knowledge Base." In *Proceedings of the Fifteenth Conference on Computational Linguistics (COLING '92)* Vol. 3, edited by Christian Boitet, 956–960. Stroudsburg, PA: Association for Computational Linguistics (ACL). <https://doi.org/10.3115/992383.992410>
- Meyer, Ingrid, Victoria Zaluski, and Kristen Mackintosh. 1997. "Metaphorical Internet Terms: A Conceptual and Structural Analysis." *Terminology* 4 (1): 1–33. <https://doi.org/10.1075/term.4.1.03mey>
- Meyer, Ralf. 1993. *Compound Comprehension in Isolation and in Context: The Contribution of Conceptual and Discourse Knowledge to the Comprehension of German Novel Noun–Noun Compounds*. Tübingen: Niemeyer. <https://doi.org/10.1515/9783111353241>
- Michalski, Ryszard S. 1991. "Concepts as Flexible and Context-Dependent Sets: The Two-Tiered View." Fairfax, VA: George Mason University (<http://www.mli.gmu.edu/papers/91-95/91-49.pdf>). Accessed 7 December 2021.
- Mickus, Timothee, Denis Paperno, Mathieu Constant, and Kees van Deemter. 2020. "What Do You Mean, BERT? Assessing BERT as a Distributional Semantics Model." *Proceedings of the Society for Computation in Linguistics* 3: Article 34. arXiv:1911.05758 [cs.CL].
- Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." In *Proceedings of the 1st International Conference on Learning Representations (ICLR)*, Scottsdale, Arizona, May 2–4, 2013, edited by Yoshua Bengio, and Yann LeCun. arXiv:1301.3781v3 [cs.CL].
- Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Distributed Representations of Words and Phrases and their Compositionality." In *Proceedings of the 26th Conference on Advances in Neural Information Processing Systems*, edited by Christopher J. C. Burges, Léon Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, 3111–3119. Red Hook, NY: Curran Associates, Inc.
- Minsky, Marvin. 1975. "A Framework for Representing Knowledge." In *The Psychology of Computer Vision*, edited by Patrick Winston, 211–277. New York: McGraw-Hill.
- Miyata, Rei, and Kyo Kageura. 2018. "Building Controlled Bilingual Terminologies for the Municipal Domain and Evaluating them Using a Coverage Estimation Approach." *Terminology* 24 (2): 149–180. <https://doi.org/10.1075/term.00017.miy>
- Moirand, Sophie. 2014. "L'événement « saisi » par la langue et la communication." *Cahiers de praxématique* 63 (<http://journals.openedition.org/praxematique/2362>). Accessed 7 December 2021. <https://doi.org/10.4000/praxematique.2362>
- Møller, Bernt. 1994. "Où va la terminologie technique française ?" *La banque des mots* 6: 69–98.
- Møller, Bernt. 1998. "À la recherche d'une terminochronie." *Meta* 43 (3): 426–438. <https://doi.org/10.7202/003655ar>
- Monière, Denis, and Dominique Labbé. 2010. "Segmentation des corpus chronologiques: 143 ans de discours gouvernemental au Québec." In *the 10th International Conference on Statistical Analysis of Textual Data*, edited by Sergio Bolasco, Isabella Chiari, and Luca Giuliano, 805–816.
- Moreno-Ortiz, Antonio. 2000. "Ontoterm®: Un sistema abierto de representación conceptual." *Procesamiento del lenguaje natural* 26: 61–64.
- Moreno-Ortiz, Antonio, and Chantal Pérez Hernández. 2002. "Ontological Semantics and Lexical Templates: Mowing the Grass from the Other Side of the Fence." In *New Perspectives on Argument Structure in Functional Grammar*, edited by Ricardo Mairal-Usón and María Jesús Pérez-Quintero, 179–212. Berlin/New York: Mouton de Gruyter, <https://doi.org/10.1515/9783110895872-005>

- Montiel-Ponsoda, Elena. 2011. *Multilingualism in Ontologies. Building Patterns and Representation Models*. Saarbruecken: Lambert Academic Publishing.
- Montiel-Ponsoda, Elena, Guadalupe Aguado-de-Cea, Asunción Gómez-Pérez, and Wim Peters. 2010. "Enriching Ontologies with Multilingual Information." In *Journal of Natural Language Engineering* 17 (3): 283–309. <https://doi.org/10.1017/S1351324910000082>
- Montiel-Ponsoda, Elena, and Guadalupe Aguado-de-Cea. 2013. "Applying the Lexical Constructional Model to Ontology Building." In *Language Processing and Grammars: The Role of Functionally Oriented Computational Models*, edited by Brian Nolan, and Carlos Perrián, 313–337. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/slcs.150.13mon>
- Montiel-Ponsoda, Elena, John McCrae, Guadalupe Aguado-de-Cea, and Jorge Gracia del Río. 2013. "Multilingual Variation in the context of Linked Data." In *Proceedings of the 10th International Conference on Terminology and Artificial Intelligence (TIA'13)*, 19–26. Villetaneuse, Francia.
- Morardo, Mikaël, and Eric Villemonte de La Clergerie. 2013. "Vers un environnement de production et de validation de ressources lexicales sémantiques." In *Actes de la 20^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2013) Atelier Sémantique Distributionnelle (SemDis 2013)*, edited by Cécile Fabre, Nabil Hathout, Philippe Muller, and Tim Van de Cruys, 167–180. Sables d'Olonne, France.
- Moreno-Ortiz, Antonio, and Chantal Pérez-Hernández. 2000. "Reusing the Mikrokosmos Ontology for Concept-Based Multilingual Terminology Databases." In *Proceedings of the 2nd International Conference on Language Resources and Evaluation. (LREC 2000)*, edited by Maria Gavrilidou, George Carayannis, Stella Markantonatou, Stelios Piperidis, and Gregory Stainhauer, 1061–1067. Athens, Greece: European Language Resources Association.
- Morgenthaler Garcia, Laura, and Carla Amorós-Negre. 2020. "Migration and Glottopolitics in the Spanish-Speaking World: Introductory Remarks." *Journal of Multilingual and Multicultural Development* 41: 1–11. <https://doi.org/10.1080/01434632.2019.1621873>
- Morgenroth, Klaus. 1990. "Le terme technique: approches théoriques et recherches didactiques appliquées à l'allemand économique d'entreprise." PhD diss., Université Paul Verlaine.
- Morin, Emmanuel. 1999. "Extraction de liens sémantiques entre termes à partir de corpus de textes techniques." PhD diss., Institut de Recherche en Informatique (IRIN), Université de Nantes.
- Morlane-Hondère, François. 2013. "Utiliser une base distributionnelle pour filtrer un dictionnaire de synonymes." In *Actes de la 20^e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2013) Atelier Sémantique Distributionnelle (SemDis 2013)*, edited by Cécile Fabre, Nabil Hathout, Philippe Muller, and Tim Van de Cruys, 112–125. Sables d'Olonne, France.
- Morrill, Charles Henry. 1997. "Language, Culture, and Society in the Central African Republic: The Emergence and Development of Sango." PhD diss., University of Michigan.
- Mortureux, Marie-Françoise. 1993. "Paradigmes désignationnels." *Semen. Revue de sémio-linguistique des textes et discours* (8). (<http://journals.openedition.org/semen/4132>). Accessed 7 December 2021. <https://doi.org/10.4000/semen.4132>
- Mortureux, Marie-Françoise. 1997. "Synonymie et co-référence dans les discours spécialisés." In *Mélanges de linguistique offerts à Rostislav Kocourek*, edited by Lise Lapierre, 185–194. Halifax: ALFA.
- Mortureux, Marie-Françoise. 2000. "Terme, (socio)terminologie et lexicologie." In *Louis Guespin, terminologie*, edited by Valérie Delavigne, and François Gaudin, 27–39. Rouen: Publications de l'Université Rouen.
- Moser, Rosemary. 1992. "Sociolinguistic Dynamics of Sango." M.A. Thesis, La Trobe University.

- Mounin, Georges. 1963. *Les problèmes théoriques de la traduction*. Paris: Gallimard.
- Mourlhon-Dallies, Florence, Florimond Rakotonelina, and Sandrine Reboul-Touré. 2004. "Les discours de l'internet : quels enjeux pour la recherche ?" *Les Carnets du Cediscor. Publication du Centre de recherches sur la didacticité des discours ordinaires* (8): 9–19. <https://doi.org/10.4000/cediscor.226>
- Muller, Charles. 1985. "4. Les langues de spécialité et les langues techniques." *Le français d'aujourd'hui*, 186–212. Paris: Éditions Klincksieck.
- Murphy, Gregory L., and Mary E. Lassaline. 1997. "Hierarchical Structure in Concepts and Basic Level of Categorization." *Knowledge, Concepts, and Categories*, edited by Koen Lamberts, and David Shanks, 93–131. Cambridge MA: MIT Press.
- Murphy, Lynne. 2003. *Semantic Relations and the Lexicon. Antonymy, Synonymy, and Other Paradigms*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486494>
- Murphy, Lynne and Anu Koskela. 2010. *Key Terms in Semantics*. New York: Bloomsbury Publishing Plc.
- Musolff, Andreas. 2016. *Political Metaphor Analysis. Discourse and Scenarios*. London: Bloomsbury.
- Myking, Johan. 2001a. "Against Prescriptivism. The 'Social-Critical' Challenge to Terminology." *Terminology Science and Research* 12 (1–2): 49–64.
- Myking, Johan. 2001b. "Sign Models in Terminology: Tendencies and Functions." *LSP and Professional Communication* 1 (2): 45–62.
- Myking, Johan. 2005. "Dichotomies: Impossible and Indispensable?" In *Proceedings of the 14th Symposium on LSP 2003*, Surrey, England, 18–22 August 2003, edited by Khurshid, Ahmad, and Margaret Rogers, 173–180. Guildford: University of Surrey.
- Myking, Johan. 2008. Motivasjon som termdanningsprinsipp: ein teoretisk diskusjon på grunnlag av norsk oljeterminologi. *Acta Wasaensia* no. 191. Vasa: Vaasan yliopisto.
- Myking, Johan. 2009. "Yet Another Taxonomy of Motivations." *Terminology Science and Research* 20: 39–53.
- Myking, Johan. 2020. "Term Formation: is There a State of the Art?" *Terminologija/ Terminology* 27: 6–30. <https://doi.org/10.35321/term27-01>

N

- Nadeau, David, and Satoshi Sekine. 2007. "A Survey of Named Entity Recognition and Classification." *Linguisticae Investigationes* 30 (1): 3–26. <https://doi.org/10.1075/li.30.1.03nad>
- Nakagawa, Hiroshi, and Tatsunori Mori. 2003. "Automatic Term Recognition based on Statistics of Compound Nouns and their Components." *Terminology* 9 (2): 201–219. <https://doi.org/10.1075/term.9.2.04nak>
- Nakos-Aupetit, Dorothy. 1977. "La recherche ponctuelle en terminologie comparée." *Meta* 22 (4): 258–263. <https://doi.org/10.7202/003688ar>
- Nakov, Preslav, and Marti Hearst. 2006. "Using Verbs to Characterize Noun-Noun Relations." In *Artificial Intelligence: Methodology, Systems, and Applications, 12th International Conference, AIMSA 2006*, edited by Jérôme Euzenat, and John Domingue. Berlin: Springer.
- Natanson, Édouard. 1981. "Synonymie des termes." *Fachsprache* 3 (2): 50–61.
- Navigli, Roberto. 2012. "A Quick Tour of Word Sense Disambiguation, Induction and Related Approaches." In *Proceedings of the 38th Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)*, edited by Mária Bielíková, Gerhard Friedrich, Georg Gottlob, and Stefan Katzenbeisser, 115–129. Berlin: Springer. https://doi.org/10.1007/978-3-642-27660-6_10

- Navigli, Roberto, and Simone Ponzetto. 2012. "BabelNet: The Automatic Construction, Evaluation and Application of a Wide-coverage Multilingual Semantic Network." *Artificial Intelligence* 193: 217–250. <https://doi.org/10.1016/j.artint.2012.07.001>
- Nazar, Rogelio. 2011. "Estudio diacrónico de la terminología especializada utilizando métodos cuantitativos: Ejemplos de aplicación a un corpus de artículos de lingüística aplicada." *Revista Signos* 44 (75): 48–67. <https://doi.org/10.4067/S0718-09342011000100005>
- Nazar, Rogelio. 2016. "Distributional Analysis Applied to Terminology Extraction. First Results in the Domain of Psychiatry in Spanish." *Terminology* 22 (2): 141–170. <https://doi.org/10.1075/term.22.2.01naz>
- Nazarenko, Adeline. 2004. "Donner accès au contenu des documents textuels. Acquisition de connaissances et analyse de corpus spécialisés." Habilitation à Diriger des Recherches, Université Sorbonne Paris Nord.
- Nedobity, Wolfgang. 1984. "Eugen Wüster und die Sprachkritiker des Wiener Kreises." *Muttersprache* 95 (1): 42–48.
- Nedobity, Wolfgang. 1985. "Terminology and Artificial Intelligence." *Knowledge Organization* 12 (1): 17–19. <https://doi.org/10.5771/0943-7444-1985-1-17>
- Nedobity, Wolfgang. 1998. "Der wissenschaftliche Briefwechsel Eugen Wüsters." In *Eugen Wüster (1898–1977), Leben und Werk – Ein österreichischer Pionier der Informationsgesellschaft – His Life and Work – An Austrian Pioneer of the Information Society*, edited by Erhard Oeser, and Christian Galinsky, 175–185. Vienna: TermNet, Infoterm.
- Neubert, Albrecht. 1972. "Theorie und Praxis für die Übersetzungswissenschaft." In *Proceedings of the Third AILA World Congress*, 482–490. Copenhagen, Denmark.
- Névéal, Aurélie, and Sylwia Ozdowska. 2006. "Terminologie médicale bilingue anglais/français: Usages clinique et législatif." *Glottopol* 8: 5–21.
- Newmark, Peter. 1981. *Approaches to Translation*. Oxford: Pergamon.
- Ney, Hermann, Ute Essen, and Reinhard Kneser. 1994. "On Structuring Probabilistic Dependencies in Stochastic Language Modelling." *Computer Speech & Language* 8 (1): 1–38. <https://doi.org/10.1006/csla.1994.1001>
- Nguyen, Thi Huyen Trang. 2013. "Analyse de l'usage du vocabulaire des matériaux de construction et description terminologique vietnamienne-française." PhD diss., Université de Rouen.
- Ngwaba, Ihuaku Chidinma. 2016. *Les termes de la gynécologie-obstétrique en igbo: enquête sur un domaine tabou dans une langue sans documents écrits*. Lyon: Université Lumière Lyon 2.
- Nida, Eugene. 1964. *Towards a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: Brill. <https://doi.org/10.1163/9789004495746>
- Nida, Eugene A., and Charles R. Taber. 1969. *The Theory and Practice of Translation*. Leiden: E. J. Brill.
- Nirenburg, Sergei, and Victor Raskin. 2004. *Ontological Semantics*. Cambridge, MA: MIT Press.
- Nistrup Madsen, Bodil, and Hanne Erdman Thomsen. 2008. "Terminological Principles used for Ontologies." In *Managing Ontologies and Lexical Resources. Proceedings of the 8th International Conference on Terminology and Knowledge Engineering, TKE 2008*, edited by Bodil Nistrup Madsen, and Hanne Erdman Thomsen, 107–122. Copenhagen: Copenhagen Business School.
- Nistrup Madsen, Bodil, Bolette Sandford Pedersen, and Hanne Erdman Thomsen. 2001. "Defining Semantic Relations for OntoQuery." In *Proceedings of the First International OntoQuery Workshop, Ontology-based interpretation of NP's*, edited by Per Anker Jensen, and Peter Skadhauge. Kolding: Department of Business Communication and Information Science, University of Southern Denmark. (<http://www.ontoquery.dk/publications/docs/Defining.doc>). Accessed 7 December 2021.

- Nord, Christiane. 1997. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome Publishing.
- Nomura, Masaaki, and Masahiko Ishii. 1988. *A List of Morpheme Combinations in Japanese Scientific Terms*. Tokyo: National Language Research Institute. (in Japanese).
- Noy, Natalya. F., and Deborah L. McGuinness. 2001. "Ontology development 101: A Guide to Creating your first Ontology." (<http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>). Accessed 7 December 2021.
- Nuopponen, Anita. 1994a. "Begræppssystem för terminologisk analys." PhD diss., University of Vaasa.
- Nuopponen, Anita. 1994b. "Wüster Revisited: On Causal Concept Relationships and Causal Concept Systems." In *Applications and Implications of Current LSP Research, Proceedings of the 9th European Symposium on LSP Vol. II*, edited by Magnar Brekke, Øivin Andersen, Trine Dahl, and Johan Myking, 532–539. Bergen: Fagbokforlaget.
- Nuopponen, Anita. 2005. "Concept Relations. An Update of a Concept Relation Classification." In *Terminology and Content Development: TKE 2005, 7th International Conference on Terminology and Knowledge Engineering*, edited by Bodil Nistrup Madsen, and Hanne Erdman Thomsen, 127–138. Litera.
- Nuopponen, Anita. 2006. "A Model for Structuring Concept Systems of Activity." In *Terminology, Standardization and Technology Transfer, Proceedings of the TSTT'2006 Conference*. Beijing, edited by Yuli Wang, Ye Wang, and Ye Tian, 83–92. Beijing: Encyclopedia of China Publishing House.
- Nuopponen, Anita. 2007. "Terminological Modelling of Processes: An Experiment." In *Indeterminacy in Terminology and LSP: Studies in Honour of Heribert Picht*, edited by Bassey E. Antia, 199–213. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.8.18nuo>
- Nuopponen, Anita. 2008. "Causal Concept Systems." In *Managing Ontologies and Lexical Resources. Proceedings of the 8th International Conference on Terminology and Knowledge Engineering, TKE 2008*, edited by Bodil Nistrup Madsen, and Hanne Erdman Thomsen, 17–28. Copenhagen: Copenhagen Business School.
- Nuopponen Anita. 2011. "Methods of Concept Analysis – Tools for Systematic Concept Analysis (Part 3 of 3)." *The LSP Journal – Language for Special Purposes, Professional Communication, Knowledge Management and Cognition* 2: 4–15.
- Nuopponen Anita. 2014. "Tangled Web of Concept Relations. Concept Relations for ISO 1087–1 and ISO 704." In *Terminology and Knowledge Engineering, TKE 2014*. Berlin (http://hal.archives-ouvertes.fr/docs/01/00/58/82/PDF/TKE2014_Nuopponen.pdf). Accessed 8 December 2021.
- Nuopponen Anita. 2016. "Satellite System as a Visualization Tool for Concept Analysis." In *Term Bases and Linguistic Linked Open Data: Proceedings of TKE 2016 the 12th International conference on Terminology and Knowledge Engineering*, edited by Hanne Erdman Thomsen, Antonio Pareja-Lora, and Bodil Nistrup Madsen, 190–200. Copenhagen: Copenhagen Business School.
- Nuopponen, Anita. 2018a. "Dimensions of Terminology Work." *Terminologija* 25: 6–22. Vilnius: Lietuvių Kalbos Institutas.
- Nuopponen, Anita. 2018b. "Terminological Concept Systems." In *Languages for Special Purposes*, edited by John Humbley, Gerhard Budin, and Christer Laurén, 453–468. Berlin/Boston: De Gruyter Mouton.

O

- Obrst, Leo. 2006. *The Ontology Spectrum and Semantic Models*. Mitre Corp. (<https://slideplayer.com/slide/12792399/>). Accessed 8 December 2021.
- Oeser, Erhard. 1998. "E. Wüster and his Impact on the Philosophy of Science." In *Eugen Wüster (1998–1977), Leben und Werk, ein österreichischer Pionier der Informationsgesellschaft*, edited by Erhard Oeser, and Christian Galinski, 105–115. Vienna, TermNet.
- Oeser, Erhard, and Gerhard Budin. 1999. "Grundlagen der Terminologiewissenschaft." In *Fachsprachen – Languages for Special Purposes, Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft – An International Handbook of Special-Language and Terminology Research*, edited by Lothar Hoffmann, and Hartwig Kalverkämper, Vol. 2, 2171–2183. Berlin: Mouton de Gruyter.
- Oeser, Erhard, and Christian Galinsky (eds). 1998. "Eugen Wüster (1898–1977), Leben und Werk – Ein österreichischer Pionier der Informationsgesellschaft – His Life and Work – An Austrian Pioneer of the Information Society." In *Proceedings of the International Conference on Professional Communication and Knowledge Transfer*. Vienna: TermNet, Infoterm.
- O'Hara, Thomas, Kavi Mahesh, and Sergei Niremburg. 1998. "Lexical Acquisition with WordNet and the Mikrokosmos Ontology." In *Proceedings of the COLING/ACL Workshop on Usage of WordNet in Natural Language Processing Systems*. Montreal, Canada. (<https://www.aclweb.org/anthology/W98-0713.pdf>). Accessed 13 July 2020.
- Otman, Gabriel. 1993. "Le talon d'Achille de l'expertise terminologique." *La banque des mots* 5: 75–87.
- Ortony, Andrew (ed.). 1993. *Metaphor and Thought*. 2nd ed. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139173865>

P

- Padó, Sebastian, and Mirella Lapata. 2007. "Dependency-based Construction of Semantic Space Models." *Computational Linguistics* 33 (2): 161–199. <https://doi.org/10.1162/coli.2007.33.2.161>
- Palomero, Vicente. 2020. "Motivacions ideològiques en la variació terminològica de la violència de gènere." *Terminàlia* 21: 29–39.
- Pan, Jeff, Guido Vetere, José Manuel Gómez-Pérez, and Honghan Wu (eds). 2017. *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Berlin/Heidelberg: Springer. <https://doi.org/10.1007/978-3-319-45654-6>
- Parizot, Anne. 2014. "Entre socioterminologie et ethnoterminologie de l'entreprise." *Textes et contextes* 9 (<http://preo.u-bourgogne.fr/textesetcontextes/index.php?id=1171>). Accessed 8 December 2021.
- Pasanen, Päivi. 2019. "Pattern-based Approach to Causality in the Field of Maritime Safety." In *New Challenges for Research on Language for Special Purposes*, edited by Ingrid Simonnæs, Øivin Andersen, and Klaus Schuber, 235–248. Berlin: Frank & Timme.
- Patterson, Karalyn, and Matthew Lambon Ralph. 2016. "The Hub-and-Spoke Hypothesis of Semantic Memory." In *Neurobiology of Language*, edited by Gregory Hickok, and Steven L. Small, 765–775. Amsterdam: Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00061-4>
- Paveau, Marie-Anne. 2014. "L'alternative quantitatif/qualitatif à l'épreuve des univers discursifs numériques." *Corela. Cognition, représentation, langage* (HS-15) (<http://corela.revues.org/3598>). Accessed 8 December 2021. <https://doi.org/10.4000/corela.3598>

- Paveau, Marie-Anne. 2015. "Ce qui s'écrit dans les univers numériques." *Itinéraires. Littérature, textes, cultures* 2014–1. <https://doi.org/10.4000/itineraires.2313>
- Pearson, Jennifer. 1998. *Terms in Context*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/scl.1>
- Pecina, Pavel. 2010. "Lexical Association Measures and Collocation Extraction." *Language Resources and Evaluation* 44 (1): 137–158. <https://doi.org/10.1007/s10579-009-9101-4>
- Pecman, Mojca. 2012a. "Étude lexicographique et discursive des collocations en vue de leur intégration dans une base de données terminologiques." *JoSTrans. The Journal of Specialised Translation* 18: 113–138.
- Pecman, Mojca. 2012b. "Tentativeness in Term Formation: A Study of Neology as a Rhetorical Device in Scientific Papers." In *Terminology* 18 (1): 27–58. <https://doi.org/10.1075/term.18.1.03pec>
- Pecman, Mojca. 2014. "Variation as a Cognitive Device: How Scientists Construct Knowledge through Term Formation." *Terminology* 20 (1): 1–24. <https://doi.org/10.1075/term.20.1.01pec>
- Pecman, Mojca. 2018. *Langue et construction de connaissances: énergie lexico-discursive et potentiel sémiotique des sciences*. Paris: L'Harmattan.
- Pecman, Mojca. 2021. "10th anniversary of the ARTES terminological and phraseological database: on the advantages for combining disciplinary with transdisciplinary language resources." In *Festschrift für Maja Bratanić/Zbornik u čast Maje Bratanić*, edited by Ivana Brač, and Ana Ostroški Anić. Zagreb, 301–324. Izdavalacko izdanje Instituta za hrvatski jezik i jezikoslovlje.
- Peñas, Anselmo, Felisa Verdejo, and Julio Gonzalo. 2001. "Corpus-based Terminology Extraction Applied to Information Access." *Special Issue: Proceedings of the Corpus Linguistics 2001, UCREL Technical Papers 13*, edited by Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie, and Shereen Khoja, 458–465. Lancashire: Lancaster University.
- Pennington, Jeffrey, Richard Socher, and Christopher D. Manning. 2014. "GloVe: Global Vectors for Word Representation." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, edited by Alessandro Moschitti, Bo Pang, and Walter Daelemans, 1532–1543. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.3115/v1/D14-1162>
- Peraldi, Sandrine. 2011. "Indétermination terminologique et multidimensionnalité dans le domaine de la chimie organique: analyse à partir d'un corpus spécialisé de langue anglaise." PhD diss., Université Paris Diderot.
- Périnet, Amandine, and Thierry Hamon. 2014. "Analyse et proposition de paramètres distributionnels adaptés aux corpus de spécialité." In *Actes des 12es Journées internationales d'analyse statistique des Données Textuelles (JADT 2014)*, edited by Emilie Née, Jean-Michel Daube, Mathieu Valette, and Serge Fleury, 507–518. Paris: Jadt. org, Analyse statistique des données textuelles.
- Périnet, Amandine, and Thierry Hamon. 2015. "Analyse distributionnelle appliquée aux textes de spécialité. Réduction de la dispersion des données par abstraction des contextes." *Traitement automatique des langues* 56 (2): 77–102.
- Perkins, David N. 1996. *Teaching for Understanding*. San Francisco CA: Exploratorium.
- Peruzzo, Katia. 2012. "Terminological Equivalence and Variation in the EU Multi-level Jurisdiction: a Case Study of Victims of Crime." PhD diss., Università degli studi di Trieste.
- Peruzzo, Katia. 2014. "Capturing Dynamism in Legal Terminology." In *Dynamics and Terminology: An Interdisciplinary Perspective on Monolingual and Multilingual Culture-bound Communication*, edited by Rita Temmerman, and Marc Van Campenhoudt, 43–60. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.16.03per>

- Péry-Woodley, Marie-Paule. 2000. "Une pragmatique à fleur de texte: approche en corpus de l'organisation textuelle." Habilitation à Diriger des Recherches, Université Toulouse le Mirail.
- Peters, Matthew E., Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. "Deep Contextualized Word Representations." In *Proceedings of the 16th North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2018)*, edited by Marilyn A. Walker, Heng Ji, and Amanda Stent, 2227–2237. Stroudsburg, PA: Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-1202>
- Petit, Gérard. 2001. "L'introuvable identité du terme technique." *Revue française de linguistique appliquée* 6 (2): 63–79. <https://doi.org/10.3917/rfla.062.0063>
- Picht, Heribert. 1981. "Terminological Activities in the Nordic Countries." In *Theoretical and Methodological Problems of Terminology. International Symposium Proceedings*, edited by Magdalena Krommer-Benz, 242–252. Vol. 6 of Infoterm Series. Munich/ London/ Paris: Sauer.
- Picht, Heribert. 1987. "Terms and their LSP Environment: LSP Phraseology." *Meta* 32 (2): 149–155. <https://doi.org/10.7202/003836ar>
- Picht, Heribert. 2005. "Knowledge Representation in Terminology." Paper presented at *Leksika, leksikografija, terminografija v russkoj, amerikanskoj i drugih kul'turach. Materialy VI Mezdunarodnoj skoly-seminara*. Ivanovo, Russia: Russian Federation.
- Picht, Heribert. 2018. "Im Anfang war." In *Terminologie und Terminologen. Festschrift für Klaus-Dirk Schmitz*, edited by Barbara Ahrens, Lisa Link, Ute B. Schilly, and Ursula Wienen, 67–74. Berlin: Frank & Timme.
- Picht, Heribert, and Jennifer Draskau. 1985. *Terminology: An Introduction*. Guildford: University of Surrey.
- Picton, Aurélie. 2009. "Diachronie en langue de spécialité, définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus: un exemple appliqué au domaine spatial." PhD diss., Université de Toulouse Le Mirail.
- Picton, Aurélie. 2011. "Picturing Short-Period Diachronic Phenomena in Specialised Corpora: A Textual Terminology Description of the Dynamics of Knowledge in Space Technologies." *Terminology* 17 (1): 134–156. <https://doi.org/10.1075/term.17.1.o8pic>
- Picton, Aurélie. 2014a. "Métalangage et sentiment linguistique des experts: regard en langue de spécialité." *Le Discours et la Langue* 6 (1): 77–90.
- Picton, Aurélie. 2014b. "The Dynamics of Terminology in Short-Term Diachrony: A Proposal for a Corpus-based Methodology to observe Knowledge Evolution." In *The Dynamics of Culture-bound Terminology in Monolingual and Multilingual Communication, Terminology and Lexicography Research and Practice*, edited by Rita Temmerman, and Marc Van Campenhoudt, 159–182. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.16.o9pic>
- Picton, Aurélie. 2018. "Terminologie outillée et diachronie: éléments de réflexion autour d'une réconciliation," *ASp* 74: 27–52. <https://doi.org/10.4000/asp.5255>
- Picton, Aurélie, Patrick Drouin, and Julie Humbert-Droz. 2021. "Exploration et visualisation de la variation terminologique en corpus spécialisés complexes: réflexions et propositions méthodologiques." In *Lexique(s) et genre(s) textuel(s) : approches sur corpus*, edited by Mathieu Mangeot, and Agnès Tutin, 99–116. Paris: Editions des archives contemporaines.
- Picton, Aurélie, and Pascaline Dury. 2017. "Diastratic Variation in Language for Specific Purposes." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francoeur, John Humbley, and Aurélie Picton, 57–80. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.o3pic>

- Picton, Aurélie, and Amélie Josselin-Leray. 2019. "A Mixed-Methods Approach to Characterize Knowledge-Rich Contexts for Specialized Translation." In *New Challenges for Research on Language for Special Purposes: Selected Proceedings from the 21st LSP Conference*, edited by Ingrid Simonæs, Øivin Andersen, and Klaus Schubert, 289–308. Berlin: Frank & Timme.
- Pierzo, Véronique. 1991. "Pour une étude socioterminologique du génie fermentaire." MA thesis, DEA de Sciences du Langage, Université de Rouen.
- Pimentel, Janine. 2012. "Criteria for the Validation of Specialized Verb Equivalents: Applications in Bilingual Terminography." PhD diss., Université de Montréal.
- Pimentel, Janine. 2015. "Adding a New Language Version to a Lexical Resource. Is it Possible to Assign Term Equivalents Semi-automatically?" *Termin@lia* 11: 20–29.
- Pinto, Sofia, Christoph Tempich, and Steffen Staab. 2004. "DILIGENT: Towards a Fine-grained Methodology for DIstributed, Loosely-controlled and EvolvInG Engineering of ONTologies." In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI04)*, 393–397. Valencia, Spain.
- Pitkänen-Heikkilä, Karina. 2015. "Adjectives as Terms." *Terminology* 21 (1): 76–101. <https://doi.org/10.1075/term.21.1.04pit>
- Poibeau, Thierry. 2005. "Parcours interprétatifs et terminologie." In *Actes du colloque Terminologie et Intelligence Artificielle*, edited by Yannick Toussaint, 12. Rouen: Université de Rouen (<https://hal.archives-ouvertes.fr/hal-00005508/document>). Accessed 8 December 2021.
- Polguère, Alain. 2012. "Propriétés sémantiques et combinatoires des quasi-prédicats sémantiques." *Scolia* 26: 131–152.
- Popa, Diana-Nicoleta. 2019. "From Lexical towards Contextualized Meaning Representation." PhD diss., Université Grenoble Alpes. HAL Id: tel-02478383.
- Poveda Villalón, María, Alba Fernández Izquierdo, and Raúl García Castro. 2019. *Linked Open Terms (LOT) Methodology (Version 1.0)*. Zenodo. <https://doi.org/10.5281/zenodo.2539305>
- Pugh, Janet. 1984. "A Contrastive Conceptual Analysis and Classification of Complex Noun Terms in English, French and Spanish with Special Reference to the Field of Data Processing." PhD diss., University of Manchester.
- Putnam, Hilary. 1975. "The Meaning of Meaning." In *Mind, Language and Reality. Philosophical Papers II*, edited by Hilary Putnam, 215–271. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511625251.014>
- Putnam, Hilary. 1975. "The Meaning of 'Meaning.'" *Minnesota Studies in the Philosophy of Science* 7: 131–193. <https://doi.org/10.1017/CBO9780511625251.014>
- Pym, Anthony. 2007. "Natural and Directional Equivalence in Theories of Translation." *Target* 19 (2): 271–294. <https://doi.org/10.1075/target.19.2.07pym>
- Pym, Anthony. 2011. "Translation Research Terms: A Tentative Glossary for Moments of Perplexity and Dispute." In *Translation Research Projects*, 75–110. Tarragona: Intercultural Studies Group.

Q

- Quillian, M. Ross. 1969. "The Teachable Language Comprehender." *Communications of the ACM*, 12 (8): 459–476. <https://doi.org/10.1145/363196.363214>
- Quiniou, Solène, Peggy, Cellier, Thierry Charnois, and Dominique Legallois. 2012. "What about Sequential Data Mining Techniques to Identify Linguistic Patterns for Stylistics?" In *Computational Linguistics and Intelligent Text Processing*, edited by Alexandra Gelbukh, 166–177. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-28604-9_14

- Quirion, Jean. 2011. "Dynamique terminologique et terminométrie: une complémentarité nécessaire pour le suivi de l'évolution de l'usage des termes et des connaissances." *Terminology* 17 (1): 113–33. <https://doi.org/10.1075/term.17.1.07qui>
- Quirion, Jean. 2013. "Facteurs sociaux de la variation terminologique dans les enquêtes d'implantation terminologique: le cas du Québec dans les années suivant l'adoption de la Charte de la Langue Française." *Debate Terminológico* 9: 47–61.
- R**
- Ramavonirina, Oliva Rahantamalala. 2011. "Langage juridique et processus de développement: le cas de Madagascar." PhD diss., Institut national des langues et civilisations orientales (Paris).
- Rastier, François. 1995. "Le terme: entre ontologie et linguistique." *La banque des mots* 7: 35–65.
- Rastier, François. 1998a. "Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage." *Langages* 129: 97–111. <https://doi.org/10.3406/lgge.1998.2149>
- Rastier, François. 1998b. "On Signs and Texts: Cognitive Science faces Interpretation." *Applied Semiotics/Sémiotique appliquée* 2(5): 195–244.
- Rastier, François. 2001. *Arts et sciences du texte*. Paris: Éditions Presses Universitaires de France. <https://doi.org/10.3917/puf.rast.2001.01>
- Rastier, François. 2003. "Parcours de production et d'interprétation: pour une conception unifiée dans une sémiotique de l'action." In *Parcours énonciatifs et parcours interprétatifs. Théories et applications*, edited by Aboubakar Ouattara, 221–242. Paris: Ophrys.
- Rastier, François. 2004. "Enjeux épistémologiques de la linguistique de corpus." *Texto!* (http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html). Accessed 8 December 2021.
- Rastier, François. 2011. *La mesure et le grain. Sémantique de corpus*. Coll. 'Lettres Numériques' 12. Paris: Honoré Champion.
- Rastier, François, Marc Cavazza, and Anne Abeillé. 1994. *Sémantique pour l'analyse*. Paris: Masson.
- Raus, Rachele. 2013. *La terminologie multilingue: la traduction des termes de l'égalité homme/femme dans le discours international*. Louvain-la-Neuve: De Boeck. <https://doi.org/10.3917/dbu.raus.2013.01>
- Rebeyrolle, Josette. 2000. "Utilisation de contextes définitoires pour l'acquisition de connaissances à partir de textes." In *Actes des Journées Francophones d'Ingénierie des Connaissances, IC 2000*, 105–114. Toulouse, France.
- Reimerink, Arianne, and Pilar León-Araúz. 2018. "Manzanilla: An Image Annotation Tool for TKB Building." In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, edited by Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, et al., 1053–1059. Paris, France: European Language Resources Association.
- Reimerink, Arianne, Pilar León-Araúz, and Pamela Faber. 2016. "Image Selection and Annotation for an Environmental Knowledge Base." *Language Resources and Evaluation* 50 (3): 1–32. <https://doi.org/10.1007/s10579-016-9345-8>
- Renwick, Adam. 2018. "Recommandations et implantation. Le cas des termes des sciences et techniques spatiales." PhD diss., Université Lyon II.
- Resche, Catherine. 2000. "An Approach to Interface Terminology: The Example of Environmental Economics in English as a Foreign Language." *Meta* 45 (4): 628–645. <https://doi.org/10.7202/003941ar>

- Resche, Catherine. 2013. "Dénominations disciplinaires et nouveaux contours d'un domaine spécialisé: le cas de la science économique." *ASp la revue du GERAS* 64: 29–50.
<https://doi.org/10.4000/asp.3825>
- Rey, Alain. 1979. *Terminologie: noms et notions*. Coll. Que sais-je ? Paris: Presses universitaires de France.
- Rey, Alain. 1995. *Essays on Terminology*. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/btl.9>
- Rey, Alain. 2011. *Dictionnaire amoureux des dictionnaires*. Paris: Plon.
- Richards, Ivor A., and Charles K. Ogden. (1923) 1989. *The Meaning of Meaning*. Orlando, Florida: Harvest/HBJ.
- Riesthuis, Gerhard (dir.). 2005. *Guidelines for Multilingual Thesauri*. IFLA (International Federation of Library Associations and Institutions) (<https://archive.ifla.org/VII/s29/pubs/Draft-multilingualthesauri.pdf>). Accessed 8 December 2021.
- Rivenc, Paul. 1971. "Lexique et langue parlée." In *La grammaire du français parlé*, edited by André Rigault, Paris: Hachette.
- Rivière, Claude. 1978. *L'analyse dynamique en sociologie*. Paris: Presses Universitaires de France.
<https://doi.org/10.3917/puf.rivie.1978.01>
- Roche, Christophe. 2007. "Le terme et le concept: fondements d'une ontoterminologie." In *Actes de la Conférence TOTh [Terminologie et ontologie] 2007*, 1–22. Chambéry: Presses Universitaires Savoie Mont-Blanc.
- Roche, Christophe. 2008. "Faut-il revisiter les principes terminologiques?" In *Actes de la Conférence TOTh [Terminologie et ontologie] 2008*, 53–72. Chambéry: Presses Universitaires Savoie Mont-Blanc.
- Roche, Christophe. 2012a. "Ontologie: entre terminologie et connaissance de spécialité." *Colloque GLAT – Terminologies: textes, discours et accès aux savoirs spécialisés*, 14–16 May 2012. Genova, Italy. (<https://hal.archives-ouvertes.fr/hal-01180282/document>). Accessed 7 December 2021.
- Roche, Christophe. 2012b. "Should Terminology Principles be re-examined?" *Proceedings of the 10th Terminology and Knowledge Engineering Conference (TKE)*, 17–32. Madrid, Spain.
- Roche, Christophe. 2012c. "Ontoterminology: How to Unify Terminology and Ontology into a Single Paradigm." In *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, 2626–2630. Istanbul, Turkey.
- Roche, Christophe. 2013. "Représentations formelles en terminologie." *Workshop TOTh [Terminologie et ontologie] 2013: Verbal and Nonverbal Representation in Terminology*. Center for Textile Research – University of Copenhagen. Copenhagen, Denmark.
- Roche, Christophe. 2015. "Ontological Definition." In *Handbook of Terminology*, edited by Henrik Kockert, and Frieda Steurs, 128–152. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/hot.1.08ont1>
- Roche, Julien. 2018. "Le Tournant ontologique de la terminologie, Informatique et langage." PhD diss., Université de Grenoble Alpes.
- Rogers, Margaret. 1997. "Synonymy and *Equivalence* in Special-Language Texts: A Case Study in German and English Texts on Genetic Engineering." In *Text Typology and Translation*, edited by Anna Trosborg, 217–245. Amsterdam/Philadelphia: John Benjamins.
<https://doi.org/10.1075/btl.26.17rog>
- Rogers, Margaret. 2000. "Genre and Terminology." In *Analysing Professional Genres*, edited by Anna Trosborg, 3–20. Pragmatics & Beyond New Series 74. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.74.03rog>

- Rogers, Margaret. 2004. "Multidimensionality in Concept Systems: A Bilingual Textual Perspective." *Terminology* 10 (2): 215–240. <https://doi.org/10.1075/term.10.2.04rog>
- Rogers, Margaret. 2007a. "Terminological Equivalence in Technical Translation: A problematic Concept?" *Fagspråk, Kommunikasjon, Kulturkunnskap* 20: 13–25.
- Rogers, Margaret. 2007b. "Lexical Chains in Technical Translation. A Case Study in Indeterminacy." In *Indeterminacy in Terminology and LSP. Terminology and Lexicography Research and Practice*, edited by Bassey E. Antia, 15–35. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.8.05rog>
- Rogers, Margaret. 2008. "Terminological Equivalence: Probability and Consistency, MuTra, LSP Translation Scenarios." In *Selected Contributions to the EU Marie Curie Conference Journal 02*, edited by Heidrun Gerzymisch-Arbogast, Gerhard Budin, and Gertrud Hofer, 101–108. Vienna, ATRC Group.
- Rogers, Margaret. 2013. "What is a 'Domain' and is this a Useful Question?" *Asp, revue du GERAS* 64: 5–16. <https://doi.org/10.4000/asp.3810>
- Rogers, Margaret. 2015. *Specialised Translation: Shedding the 'Non-literary' Tag*. Berlin: Springer. <https://doi.org/10.1057/9781137478412>
- Rogers, Margaret, and Sue Ellen Wright. 2006. "Approaches to Terminological Theories: the Anglo-Saxon Approach." In *Modern Approaches to Terminological Theories and Applications*, edited by Heribert Picht, 107–134. Bern: Peter Lang.
- Rohrer, Tim. 2006. "Three Dogmas of Embodiment: Cognitive Linguistics as a Cognitive Science." In *Cognitive Linguistics: Current Applications and Future Perspectives*, edited by Gitte Kristiansen, Michel Achard, René Dirven, and Francisco Ruiz de Mendoza, 119–146. Berlin/New York: Mouton de Gruyter.
- Rohrer, Tim. 2007. "Embodiment and Experientialism." In *The Oxford Handbook of Cognitive Linguistics*, edited by Dirk Geeraerts, and Hubert Cuyckens, 25–47. Oxford: Oxford University Press.
- Rojas-García, Juan, and Melania Cabezas-García. 2019. "Use of Knowledge Patterns for the Evaluation of Semi-automatically Induced Semantic Clusters." In *New Challenges for Research on Language for Special Purposes*, edited by Ingrid Simonnaes, Øivin Andersen, and Klaus Schubert, 121–140. Serie Forum für Fachsprachen-Forschung, vol. 154. Berlin: Frank & Timme.
- Rojas-García, Juan, and Pamela Faber. 2019. "Evaluation of Distributional Semantic Models for the Extraction of Semantic Relations for Named Rivers from a Small Specialized Corpus." *Procesamiento del Lenguaje Natural* 63: 51–58.
- Romary, Laurent, Isabelle Kramer, Susanne Salmon-Alt and Joseph Roumier. 2006. "Gestion de données terminologiques : principes, modèles, méthodes." In *Terminologie et accès à l'information*, edited by Widad Mustafa El Hadi. Paris: Hermès. (<https://hal.archives-ouvertes.fr/hal-00096910/document>). Accessed 8 December 2021.
- Rondeau, Guy. 1984. *Introduction à la terminologie*. Chicoutimi: Gaëtan Morin.
- Rosch, Eleanor. 1978. "Principles of Categorization." In *Cognition and Categorization*, edited by Eleanor Rosch, and Barbara Lloyd, 27–48. Hillsdale, NJ: Erlbaum.
- Rosch, Eleanor, and Carolyn Mervis. 1975. "Family Resemblances: Studies in the Internal Structure of Categories." *Cognitive Psychology* 7 (4): 573–605. [https://doi.org/10.1016/0010-0285\(75\)90024-9](https://doi.org/10.1016/0010-0285(75)90024-9)
- Rossi, Micaela. 2017. "Terminological Metaphors and the Nomadism of Specialized Terms: Some Observations on Intralinguistic and Interlinguistic Variation." In *Multiple Perspectives on Terminological Variation*, edited by Patrick Drouin, Aline Francœur, John Humbley, and Aurelie Picton, 181–211. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.18.08ros>

- Rossi, Micaela. 2018. "Les métaphores terminologiques dans les sciences et leur traduction dans les textes et discours, entre invention et vulgarisation." In *Terminology & Discourse / Terminologie et discours*, edited by Jana Altmanova, Maria Centrella, and Katherine E. Russo, 179–196. Bern: Peter Lang.
- Rotaru, Armand Stefan and Gabriella Vigliocco. 2019. "Modelling Semantics by Integrating Linguistic, Visual and Affective Information." In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society (CogSci): Creativity + Cognition + Computation*, edited by Ashok K. Goel, Colleen M. Seifert, and Christian Freksa, 2681–2687. Red Hook, NY: Curran Associates, Inc.
- Roulon-Doko, Paulette. 1996. *Conception de l'espace et du temps chez les Gbaya de Centrafrique*. Paris: L'Harmattan.
- Roulon-Doko, Paulette. 1998. *Chasse, culture et cueillette chez les Gbaya de Centrafrique*. Paris: Editions L'Harmattan.
- Ruiz de Mendoza, Francisco José. 1997. "Metaphor, Metonymy and Conceptual Interaction." *Atlantis* 19 (1): 281–295.
- Ruiz de Mendoza, Francisco José, and Alicia Galera. 2014. *Cognitive Modeling. A Linguistics Perspective*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/hcp.45>
- Ruiz de Mendoza, Francisco José, and Lorena Pérez. 2011. "The Contemporary Theory of Metaphor: Myths, Developments and Challenges." *Metaphor and Symbol* 26 (3): 161–185. <https://doi.org/10.1080/10926488.2011.583189>
- Ryan, Richard. 2018. "Controlled Language." In *Languages for Special Purposes: An International Handbook*, edited by John Humbley, Gerhards Budin, and Laurén Christer, 289–306. Berlin/Boston: De Gruyter Mouton.

S

- Sageder, Dagmar. 2010. "Terminology Today: A Science, an Art or a Practice? Some Aspects on Terminology and its Development." *Brno Studies in English* 36 (1): 123–134.
- Sager, Juan C. 1990. *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.44>
- Sager, Juan C. 1994a. "What's Wrong with Terminology Work and Terminology Science?" *Terminology* 1 (2): 375–381. <https://doi.org/10.1075/term.1.2.09sag>
- Sager, Juan C. 1994b. "Reflections on Terminological Translation Equivalents." *Lebende Sprachen* 39 (2): 55–57. <https://doi.org/10.1515/les.1994.39.2.55>
- Sager, Juan C. 1997. "Term Formation." In *Handbook of Terminology Management*, edited by Sue Ellen Wright, and Gerhard Budin, 25–41. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.htm1.06sag>
- Sager, Juan C. 1998. "Search for a Foundation towards a Theory of the Term." *Terminology* 5 (1): 41–57. <https://doi.org/10.1075/term.5.1.05sag>
- Sager, Juan C. 2003. "Das Wort der Welt, schaubildlich und terminologisch dargestellt. The Wording of the World presented Graphically and Terminologically." *Terminology* 9 (2): 269–297. <https://doi.org/10.1075/term.9.2.08sag>
- Sahlgren, Magnus. 2005. "An Introduction to Random Indexing." In *Proceedings of the Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering (TKE 2005)*, edited by Bodil N. Madsen, and Hanne E. Thomsen. Copenhagen, Denmark (<https://folk.idi.ntnu.no/gamback/teaching/TDT4138/sahlgren05.pdf>). Accessed 8 December 2021.

- Sahlgren, Magnus. 2006. "The Word-Space Model." PhD diss., Stockholm University.
- Sahlgren, Magnus. 2008. "The Distributional Hypothesis." *Rivista di Linguistica* 20 (1): 33–53.
- Sahlgren, Magnus, and Jussi Karlgren. 2009. "Terminology Mining in Social Media." In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)*, edited by David Wai-Lok Cheung, Il-Yeol Song, Wesley W. Chu, Xiaohua Hu, and Jimmy J. Lin, 405–414. New York: Association for Computing Machinery.
<https://doi.org/10.1145/1645953.1646006>
- Sahlgren, Magnus, and Alessandro Lenci. 2016. "The Effects of Data Size and Frequency Range on Distributional Semantic Models." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, edited by Jian Su, Kevin Duh, and Xavier Carreras, 975–980. Stroudsburg, PA: Association for Computational Linguistics.
<https://doi.org/10.18653/v1/D16-1099>
- Salager-Meyer, Françoise, María Ángeles Alcaraz Ariza, and Nahirana Zambrano. 2003. "The Scimitar, the Dagger and the Glove: Intercultural Differences in the Rhetoric of Criticism in Spanish, French and English Medical Discourse (1930–1995)." *English for Specific Purposes* 22 (3): 223–247. [https://doi.org/10.1016/S0889-4906\(02\)00019-4](https://doi.org/10.1016/S0889-4906(02)00019-4)
- Samain, Didier. 2008. "Eugen Wüster. De l'esperanto à la terminologie." *Cultures et lexicographie*, 2008, Germany, 279–296. {halshs-00529326}. Accessed 8 December 2021.
- Sandoval, Kristopher. 2020. "The Difference between Tight Coupling and Loose Coupling." (<https://nordicapis.com/the-difference-between-tight-coupling-and-loose-coupling/>). Accessed 8 December 2021.
- Sandrini, Peter. 1996. "Comparative Analysis of Legal Terms: Equivalence Revisited." In *Proceedings of the Fourth International Congress on Terminology and Knowledge Engineering (TKE '96)*, edited by Christian Galinski, and Klaus-Dirk Schmitz, 342–351. Frankfurt: Indeks.
- San Martín, Antonio. 2016. "La representación de la variación contextual mediante definiciones terminológicas flexibles." PhD diss. University of Granada.
- San Martín, Antonio, Melania Cabezas-García, Miriam Buendía-Castro, Beatriz Sánchez-Cárdenas, Pilar León-Araúz, and Pamela Faber. 2017. "Recent Advances in EcoLexicon." *Dictionaries: Journal of the Dictionary Society of North America* 38 (1): 96–115.
<https://doi.org/10.1353/dic.2017.0004>
- San Martín, Antonio, Melania Cabezas-García, Miriam Buendía-Castro, Beatriz Sánchez-Cárdenas, Pilar León-Araúz, Arianne Reimerink, and Pamela Faber. 2020. "Presente y futuro de la base de conocimiento terminológica EcoLexicon." *Onomázein* 49: 174–202.
<https://doi.org/10.7764/onomazein.49.09>
- San Martín, Antonio, Catherine Trekker, and Pilar León-Araúz. 2020. "Extraction of Hyponymic Relations in French with Knowledge-pattern-based Word Sketches." In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, edited by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, et al., 5955–5963. Marseille: European Language Resources Association.
- Sanz-Vicente, Lara. 2012. "Approaching Secondary Term Formation through the Analysis of Multiword Units: An English–Spanish Contrastive Study." *Terminology* 18 (1): 105–127.
<https://doi.org/10.1075/term.18.1.06san>
- Sapir, Edward. 1961. *Culture, Language and Personality. Selected Essays*, edited by David G. Mandelbaum. Berkeley/Los Angeles: University of California Press.
- Šarčević, Susan. 1988. "The Challenge of Legal Lexicography: Implications for Bilingual and Multilingual Dictionaries." In *Zurilex '86 Proceedings*, edited by Mary Snell-Hornby, 307–314. Tübingen: Francke.

- Šarčević, Susan. 1989. "Conceptual Dictionaries for Translation in the Field of Law." *International Journal of Lexicography* 2 (4): 277–293. <https://doi.org/10.1093/ijl/2.4.277>
- Šarčević, Susan. 1997. *New Approach to Legal Translation*. The Hague: Kluwer Law International.
- Saussure, Ferdinand de. (1916, 1959) 1975. *Cours de linguistique générale*. Paris: Payot.
- Schaetzen, Caroline de (ed.). 1989. *Terminologie diachronique. Proceedings of the First International Conference on Diachronic Terminology*. Paris: CILF.
- Schlomann Lowe, Elizabeth, and Sue Ellen Wright. 2006. "The Life and Works of Alfred Schlomann: Terminology Theory and Globalization." In *Modern Approaches to Terminological Theory and Applications*, edited by Herbert Picht, 153–161. Bern: Peter Lang.
- Schmidt, Thomas. 2009. "The Kicktionary – A Multilingual Lexical Resource of Football Language." In *Multilingual FrameNets in Computational Lexicography, Methods and Applications*, edited by Hans C. Boas, 101–134. Berlin/New York: Mouton de Gruyter.
- Schnitzer, Johannes. 2006. "La diversidad diatópica en terminología: el caso de la terminología económica española." In *La terminología en el siglo XXI: Actas del IX Simposio Iberoamericano de Terminología RITERM04*, edited by María Teresa Cabré, Rosa Estopà, and Carles Tebé, 595–604. Barcelona: Documenta Universitaria.
- Schumann, Anne-Kathrin. 2012. "Towards the Automated Enrichment of Multilingual Terminology Databases with Knowledge-Rich Contexts – Experiments with Russian EuroTermBank Data." In *Proceedings of CHAT 2012: The 2nd Workshop on the Creation; Harmonization and Application of Terminology Resources*, edited by Tatiana Gornostay, 27–34. Linköping: Linköping Electronic Conference Proceedings.
- Schurz, Gerhard, and Ioannis Votsis. 2015. "Reconstruction Scientific Theory Change by Means of Frames." In *Meaning, Frames, and Conceptual Representation*, edited by Tomas Gamerschlag, Doris Gerland, Rainer Osswald, and Wiebke Petersen, 93–109. Düsseldorf: Düsseldorf University Press.
- Schütze, Hinrich. 1998. "Automatic Word Sense Discrimination." *Computational Linguistics* 24 (1): 97–123.
- Seghezzi, Natalia. 2013. "Variación terminológica: de la escritura a la oralidad." *Debate Terminológico* 9: 38–46.
- Séguéla, Patrick. 2001. "Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques." PhD diss., Université Toulouse III.
- Shapiro, Lawrence. 2004. *The Mind Incarnate*. Cambridge, MA: MIT Press.
- Sharifian, Farzad. 2008. "Distributed, Emergent Cultural Cognition, Conceptualisation, and Language." In *Body, Language and Mind. Vol. II. Interrelations between Biology, Linguistics and Culture*, edited by Roslyn M. Frank, René Dirven, Tom Ziemke, and Enrique Bernárdez, 109–136. Tübingen: Springer.
- Sharifian, Farzad. 2017. "Cultural Linguistics and Linguistic Relativity." *Language Sciences* 59: 83–92. <https://doi.org/10.1016/j.langsci.2016.06.002>
- Shaw, Philip. 2006. "Text Purpose, Text Effects, Test Power English for Specific Purposes." *The European English Messenger* 15 (2): 39–43.
- Shreve, Gregory. 1995. "SGML Representation of Concept Systems: Identifying, Tagging and Retrieving Term-Concept Structures in Textual Context." In *Standardizing and Harmonizing Terminology: Theory and Practice*, edited by Sue-Ellen Wright, and Richard A. Strehlow, 157–168. West Conshohocken, PA: ASTM International. <https://doi.org/10.1520/STP13753S>
- Sierra, Gerardo, Rodrigo Alarcón, César Aguilar, and Carme Bach. 2008. "Definitional Verbal Patterns for Semantic Relation Extraction." *Terminology* 14 (1): 74–98. <https://doi.org/10.1075/term.14.1.05sie>
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

- Sinclair, John. 2005. "Corpus and Text – Basic Principles." In *Developing Linguistic Corpora: A Guide to Good Practice*, edited by Martin Wynn (http://icar.cnrs.fr/ecole_thematique/cont-aci/documents/Baude/wynne.pdf). Accessed 8 December 2021.
- Sinha, Chris. 1999. "Grounding, Mapping and Acts of Meaning." In *Cognitive linguistics: Foundations, Scope and Methodology*, edited by Theo Janssen, and Gisela Redeker, 223–255. Berlin: Mouton de Gruyter. <https://doi.org/10.1515/9783110803464.223>
- Sivilotti, Lucas. 2019. "La médiation auprès des étudiants atteints de cancer ou de maladies rares: analyse et compréhension des enjeux. Étude en Nouvelle-Aquitaine autour de la trajectoire étudiante et des interactions plurielles." PhD diss., Université de Bordeaux.
- Skopnik-Chicago, Marianne, Arlette Bassaber, and Pablo A. Lizana. 2020. "Análisis de los Textos Escolares de Ciencias Naturales y Biología desde la Perspectiva de la Terminología Anatómica Internacional: Sistema Cardiovascular." *International Journal of Morphology* 38 (6): 1751–1759. <https://doi.org/10.4067/S0717-95022020000601751>
- Skuce, Douglas. 1989. "A Generic Knowledge Acquisition Environment Integrating Natural Language and Logic." In *Proceedings of the JCAI Workshop on Knowledge Acquisition*. Detroit, Michigan.
- Skuce, Douglas, and Ingrid Meyer. 1990. "Computer-Assisted Concept Analysis: An Essential Component of a Terminologist's Workstation." In *TKE '90: Terminology and Knowledge Engineering*, Vol. I, 18. 7–199. Frankfurt: INDEKS Verlag.
- Skuce, Douglas, and Ingrid Meyer. 1991. "Terminology and Knowledge Engineering: Exploring a Symbiotic Relationship." In *Proceedings of the 6th International Workshop on Knowledge Acquisition for Knowledge-Based Systems*, 29/1–29/21. Banff, Canada.
- Skuce, Douglas, and Ira Monarch. 1990. "Ontological Issues in Knowledge Base Design: Some Problems and Suggestions." In *Proceedings of the 5th Workshop on Knowledge Acquisition for Knowledge-Based Systems*. Banff, Canada.
- Skuce, Douglas, S. Wang, and Yves Beauvillé. 1989. "A Generic Knowledge Acquisition Environment for Conceptual and Ontological Analysis." In *Proceedings of the Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Canada.
- Slodzian, Monique. 1986. "Présentation." *La terminologie en URSS: aperçu sur ses développements en Europe Centrale*. Slovo, Publications Langues'O 8: 3–8.
- Slodzian, Monique. 1994. "Comprendre pour communiquer et soigner." In *Comprendre pour communiquer et soigner, langues, informatique et santé oculaire en Afrique*, edited by Robert Chaudenson, and Monique Slodzian, 45–54. Paris: Didier Erudition.
- Slodzian, Monique. 1995. "Comment revisiter la doctrine terminologique aujourd'hui ?" *La banque des mots* 7: 11–18.
- Slodzian, Monique. 2000. "L'émergence d'une terminologie textuelle et le retour du sens." In *Le sens en terminologie*, edited by Henri Béjoint, and Philippe Thoirion, 61–85. Lyon: Presses Universitaires de Lyon.
- Smith, Barry. 2004. "Beyond Concepts: Ontology as Reality Representation." In *Proceedings of the 3rd International Conference on Formal Ontology in Informatic Systems (FOIS04)*, 73–84. Amsterdam: IOS Press.
- Smith, Barry, Werner Cuesters, and Rita Temmerman. 2005. "Wüsteria." *Studies in Health Technology and Informatics* 116: 647–652.
- Smith, Viktor, Ditte Green-Petersen, Peter Møgelvang-Hansen, Rune H. Christensen, Françoise Qvistgaard, and Grethe Hyldig. 2013. "What's (in) a Real Smoothie. A Division of Linguistic Labour in Consumers' Acceptance of Name-Product Combinations?" *Appetite* 63: 129–140. <https://doi.org/10.1016/j.appet.2012.10.020>

- Snell-Hornby, Mary. 1988. *Translation Studies. An Integrated Approach*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.38>
- Soler, Victoria, and Amparo Alcina. 2008. "Patrones léxicos para la extracción de conceptos vinculados por la relación parte-todo en español." *Terminology* 14 (1): 99–123. <https://doi.org/10.1075/term.14.1.06sol>
- Staab, Steffen, Hans Peter Schnurr, Rudi Studer, and York Sure. 2001. "Knowledge Processes and Ontologies." *IEEE Intelligent Systems, Special Issue on Knowledge Management* 16 (1): 26–34. <https://doi.org/10.1109/5254.912382>
- Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Štekauer, Pavol. 1998. *An Onomasiological Theory of English Word-Formation*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/sfsl.46>
- Štekauer, Pavol, Salvador Valera, and Livia Körtvélyessy. 2012. *Word-Formation in the World's Languages*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511895005>
- Steen, Gerard. 2007. *Finding Metaphor in Grammar and Usage*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/celcr.10>
- Steen, Gerard. 2008. "The Paradox of Metaphor: Why We Need a Three-Dimensional Model of Metaphor." *Metaphor and Symbol* 23 (4): 213–241. <https://doi.org/10.1080/10926480802426753>
- Steen, Gerard. 2017. "Deliberate Metaphor Theory: Basic Assumptions, Main Tenets, Urgent Issues." *Intercultural Pragmatics* 14 (1): 1–24. <https://doi.org/10.1515/ip-2017-0001>
- Studer, Rudi, Richard Benjamins, and Dieter Fensel. 1998. "Knowledge Engineering: Principles and Methods." *Data & Knowledge Engineering* 25 (1–2): 161–198. [https://doi.org/10.1016/S0169-023X\(97\)00056-6](https://doi.org/10.1016/S0169-023X(97)00056-6)
- Suárez-Figueroa, Mari Carmen, Asunción Gómez-Pérez, and Boris Villazón-Terrazas. 2009. "How to Write and Use the Ontology Requirements Specification Document." In *On the Move to Meaningful Internet Systems: OTM 2009*, edited by Robert Meersman, Tharam Dillon, and Pilar Herrero, 966–982. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-05151-7_16
- Suárez-Figueroa, Mari Carmen, Asunción Gómez-Pérez, and Mariano Fernández-López. 2012. "The NeOn Methodology for Ontology Engineering." In *Ontology Engineering in a Networked World*, 9–34. Berlin/Heidelberg: Springer. https://doi.org/10.1007/978-3-642-24794-1_2
- Suárez-Toste, Ernesto. 2017. "Babel of the Senses: On the Role of the Metaphor and Synesthesia in Wine Reviews." *Terminology* 23 (1): 89–112. <https://doi.org/10.1075/term.23.1.04sua>
- Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Sweetser, Eve, Oana David, and Elise Stickles. 2019. "MetaNet: Automated Metaphor Identification across Languages and Domains." In *Metaphor and Metonymy in the Digital Age*, edited by Marianna Bolognesi, Mario Brdar, and Kristina Despot, 23–47. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/milcc.8.02swe>

T

- Taconnat, Laurence. 2012. "Fonctionnement et dysfonctionnement de la mémoire humaine." *Le journal des psychologues* 297: 18–23. <https://doi.org/10.3917/jdp.297.0018>
- Talmy, Leonard. 1996. "Fictive Motion in Language and 'Ception.'" In *Language and Space*, edited by Paul Bloom, Mary Peterson, Lynn Nadel, and Merrill Garrett, 211–276. Cambridge, MA: MIT Press.

- Talmy, Leonard. 2000. *Toward a Cognitive Semantics* (Vol. I). Cambridge, MA: MIT Press.
- Tanguy, Ludovic. 2013. "La ruée linguistique vers le Web." *Texto!* 18 (4) (<http://www.revue-texto.net/index.php?id=3335>). Accessed 8 December 2021.
- Tanguy, Ludovic, Pauline Brunet, and Olivier Ferret. 2020. "Extrinsic Evaluation of French Dependency Parsers on a Specialised Corpus: Comparison of Distributional Thesauri." In *Proceedings of the 12th Language Resources and Evaluation, LREC 2020*, edited by Nicoletta Calzolari, Frédéric B  chet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, et al., 5820–5828. Paris: ELRA.
- Tanguy, Ludovic, C  cile Fabre, and Yoann Bard. 2020. "Impact de la structure logique des documents sur les mod  les distributionnels: exp  rimentations sur le corpus TALN." In *Actes de la 27   Conf  rence sur le Traitement Automatique des Langues Naturelles (TALN 2020)* (vol.2), edited by Christophe Benzitoun, Chlo   Braud, Laurine Huber, David Langlois, Slim Ouni, Sylvain Pogodalla, and St  phane Schneider, 122–135. ATALA and AFCP.
- Tanguy, Ludovic, Franck Sajous, and Nabil Hathout. 2015. "Evaluation sur mesure de mod  les distributionnels sur un corpus sp  cialis  : comparaison des approches par contextes syntaxiques et par fen  tres graphiques." *Traitement automatique des langues* 56 (2): 103–127.
- Taylor, James, and Elizabeth Van Every. 2000. *The Emergent Organization: Communication as its Site and Surface*. Mahwah, NJ: Lawrence Erlbaum.
- Teb  , Carles. 2005. *La representaci   conceptual en terminologia. l'atribuci   tem  tica en els bancs de dades terminol  giques*. Barcelona: Institut Universitari de Ling  stica Aplicada.
- Temmerman, Rita. 1997. "Questioning the Univocity Ideal. The Difference Between Socio-cognitive Terminology and Traditional Terminology." *Hermes, Journal of Linguistics* 18: 51–90.
- Temmerman, Rita. 2000. *Towards New Ways of Terminology Description. The Sociocognitive Approach*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.3>
- Temmerman, Rita. 2002. "Metaphorical Models and the Translation of Scientific Texts." *Linguistica Antverpiensia* 1: 211–226.
- Temmerman, Rita. 2007. "Approaches to Terminology. Now that the Dust has Settled..." *SYNAPS* 20: 27–36.
- Temmerman, Rita. 2008. "Sociocultural Situatedness of Terminology in the Life Sciences: The History of Splicing." In *Body, Language and Mind (Vol. II)*, edited by Tom Ziemke, Jordan Zlatev, and Roslyn Frank, 327–360. Berlin: Walter de Gruyter.
- Temmerman, Rita. 2013. "Primary and Secondary Term Creation and the Process of Understanding." In *Application of Cognitive Terminology Theories in Terminology Management*. Zagreb, Croatia (<https://studylib.net/doc/8899263/primary-and-secondary-term-creation-and-the-process-of>). Accessed 8 December 2021.
- Temmerman, Rita. 2014. "How does Molecular Biology Terminology Develop in English and French?" In *Traduire aux confins du lexique: Les nouveaux terrains de la terminologie*, edited by Marc Van Campenhoudt, Nathalie Lemaire, and Rita Temmerman, 129–154. Coll. Cahiers de linguistique Vol. 40/2. Brussels: EME Editions.
- Temmerman, Rita. 2016. "Translation and the Dynamics of Understanding Words and Terms in Contexts." In *Meaning in Translation: Illusion of Precision*, edited by Larisa Ilynska, and Marina Platonova, 139–160. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Temmerman, Rita. 2017. "Verbalizing Sensory Experience for Marketing Success: The Case of the Wine Descriptor 'minerality' and the Product Name 'smoothie.'" *Terminology* 23 (1): 132–154. <https://doi.org/10.1075/term.23.1.o6tem>
- Temmerman, Rita. 2018. "European Union Multilingual Primary Term Creation and the Impact of its Neologisms on National Adaptations." *Parall  les* 30 (1) 8–20.

- Temmerman, Rita. Forthcoming. "Terminological Metaphors: Framing for Better and for Worse ...". In *On Metaphor*, edited by Michele Prandi, and Micaela Rossi. London: Routledge.
- Temmerman, Rita, and Koen Kerremans. 2003. "Termontography: Ontology Building and the Sociocognitive Approach to Terminology Description." In *Proceedings of CIL17*, edited by Eva Hajičová, Anna Kotěšovcová, and Jiří Mírovský. Prague, Czech Republic: Matfyzpress.
- Temmerman, Rita, and Marc Van Campenhoudt (eds). 2011. "The Dynamics of Terms in Specialized Communication: An Interdisciplinary Perspective." Special Issue of *Terminology* 17 (1): 1–8. <https://doi.org/10.1075/term.17.1.01tem>
- Tercedor, Maribel. 1999. "Rutas de metaforización y traducción especializada: una aproximación cognitiva." *Sendeban* 10–11: 249–260.
- Tercedor, Maribel. 2004. "Descripción y variación de la representación terminológica: el caso de la dimensión tipos de cáncer." In *Investigar en terminología*, edited by Pamela Faber, and Catalina Jiménez, 199–214. Granada: Comares.
- Tercedor, Maribel. 2011. "The Cognitive Dynamics of Terminological Variation." *Terminology* 17 (2): 181–197. <https://doi.org/10.1075/term.17.2.01ter>
- Tercedor, Maribel. 2013. "Una perspectiva situada de la variación denominativa." *Debate Terminológico* 9: 81–88.
- Tercedor, Maribel, and Clara Inés López. 2012. "Access to Health in an Intercultural Setting: The Role of Corpora and Images in Grasping Term Variation." *Linguistica Antverpiensia, New Series – Themes in Translation Studies* 11: 247–268.
- Tesnière, Lucien. 1959. *Éléments de Syntaxe Structurale*. Paris: Éditions Klincksieck.
- Teubert, Wolfgang. 2005. "My Version of Corpus Linguistics." *International Journal of Corpus Linguistics* 10 (1): 1–13. <https://doi.org/10.1075/ijcl.10.1.01teu>
- Thayer, Charles. 1979. "Biological Bulldozers and the Evolution of Marine Benthic Communities." *Science* 203: 458–461. <https://doi.org/10.1126/science.203.4379.458>
- Thibaudet, Albert. 1936. *Histoire de la littérature française de 1789 à nos jours*. Paris (<https://gallica.bnf.fr/ark:/12148/bpt6k6276700w/f568.item#>). Accessed 8 December 2021.
- Thoiron, Philippe, Jean Iwaz, and Nadine Zaouche. 1997. "Étude d'implantation des arrêtés de terminologie. Domaines : santé et médecine." In *La mesure des mots, cinq études d'implantation terminologique*, edited by Loic Depecker, 47–95. Rouen: Publications de l'Université de Rouen.
- Thornbury, Scott. 1999. *How to Teach Grammar*. Essex: Pearson Education Limited.
- Tourneux, Henry. 2006. *La communication technique en langues africaines, l'exemple de la lutte contre les ravageurs du cotonnier (Burkina Faso / Cameroun)*. Paris: Karthala.
- Tourneux, Henry, (ed). 2008. *Langues, cultures et développement en Afrique*. Paris: Karthala.
- Tourneux, Henry, Yaya Dairou, Abdoulaye Boubakary. 2017. *Dictionnaire peul encyclopédique de la nature (faune/flore), de l'agriculture, de l'élevage et des usages en pharmacopée (Diamaré, Cameroun), : suivi d'un index médicinal et d'un index français-fulfulde*. Yaoundé: Cerdotola.
- Tran, Tuan Duc, Anita Burgun, and Nicolas Garcelon. 2003. "Acquisition semi-automatique de terminologie bilingue en biologie moléculaire à partir des corpus comparables." In *Terminologie et Intelligence Artificielle (TIA'03)*, 166–175. Strasbourg, France.
- Trojar, Mitja. 2017. "Wüster's View of Terminology." *Slovenski jezik- Slovene Linguistic Studies* 11: 55–85.
- Trosborg, Anna. 2000. *Analysing Professional Genres. Pragmatics & Beyond New Series 74*. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/pbns.74>
- Turner, Mark, and Gille Fauconnier. 1995. "Conceptual Integration and Formal Expression." *Metaphor and Symbolic Activity* 10: 183–204. https://doi.org/10.1207/s15327868ms1003_3

- Turney, Peter D., and Patrick Pantel. 2010. "From Frequency to Meaning: Vector Space Models of Semantics." *Journal of Artificial Intelligence Research* 37: 141–188. <https://doi.org/10.1613/jair.2934>
- Tyler, Andrea, and Vyvyan Evans. 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511486517>

U

- Uffmann, Christian. 2015. "Loanword adaptation." In *The Oxford Handbook of Historical Phonology*, edited by Patrick Honeybone, and Joseph Salmons, Oxford: Oxford University Press (<https://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199232819.001.0001/oxfordhb-9780199232819-e-019>). Accessed 8 December 2021.
- Ureña, José Manuel. 2011. "Metaphor in Specialised Language: An English-Spanish Comparative Study in Marine Biology." PhD diss., University of Granada.
- Ureña, José Manuel. 2016. "Refining the Understanding of Novel Metaphor in Specialised Language Discourse." *Terminology* 22 (1): 1–29. <https://doi.org/10.1075/term.22.1.01ure>
- Ureña, José Manuel, and Pamela Faber. 2010. "Reviewing Imagery in Resemblance and Non-Resemblance Metaphors." *Cognitive Linguistics* 21: 123–149. <https://doi.org/10.1515/cogl.2010.004>
- Ureña, José Manuel, and Pamela Faber. 2011. "Strategies for the Semi-Automatic Retrieval of Metaphorical Terms." *Metaphor and Symbol* 26 (1): 23–52. <https://doi.org/10.1080/10926488.2011.535415>
- Ureña, José Manuel, and Pamela Faber. 2014. "A Cognitive Sociolinguistic Approach to Metaphor and Denominative Variation. A Case Study of Marine Biology Terms." *Review of Cognitive Linguistics* 12 (1): 193–222. <https://doi.org/10.1075/rcl.12.1.07ure>
- Ureña, José Manuel, and Maribel Tercedor. 2011. "Situated Metaphor in Scientific Discourse." *Languages in Contrast* 11 (2): 216–240. <https://doi.org/10.1075/lic.11.2.04ure>
- Uschold, Michael. 1998. "Knowledge Level Modelling: Concepts and Terminology." In *The Knowledge Engineering Review* 13 (1): 5–29. <https://doi.org/10.1017/S0269888998001040>
- Uschold, Michael, and Martin King. 1995. "Towards a Methodology for Building Ontologies." In *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing at IJCAI95*, 15–30. Edinburgh, UK (<http://www.aiai.ed.ac.uk/publications/documents/1995/95-ont-ijcai95-ont-method.pdf>). Accessed 8 December 2021.

V

- Valeontis, Kostas, and Elena Mantzari. 2006. "Linguistic Dimension of Terminology: Principles and Methods of Term Formation." In *Proceedings of the 1st Athens International Conference on Translation and Interpretation Translation: Between Art and Social Science*, 1–20. Athens, Greece.
- Van Campenhoudt, Marc. 1997. "Evaluation des terminographies multilingues : le dictionnaire nautique du capitaine Paasch face au dictionnaire aéronautique de Schlomann." In *Les dictionnaires spécialisés et l'analyse de la valeur*, edited by Ad Hermans, 75–115. Leuven: Peeters.
- Van Campenhoudt, Marc. 1998. "Maille ou maillon: quand les terminographes négligent l'évolution de l'usage." In *La mémoire des mots*, edited by André Clas, Salah Mejri, and Taïeb Baccouche, 251–272. Tunis: AUPELF-UREF.
- Van Campenhoudt, Marc. 2001. "Pour une approche sémantique du terme et de ses équivalents." *International Journal of Lexicography* 14 (3): 181–209. <https://doi.org/10.1093/ijl/14.3.181>

- Van Campenhoudt, Marc. 2006. "Que nous reste-t-il d'Eugen Wüster ? Paper session presented at *Eugen Wüster et la terminologie de l'École de Vienne* (03-04/02/2006: Université de Paris 7). (<https://dipot.ulb.ac.be/dspace/bitstream/2013/238799/3/wuster.pdf>). Accessed on 8 December 2021.
- Van der Yeugt, Michel. 2018. "Charting the Diachronic Dimension of Specialised Languages: Epistemological Challenges and Pedagogical Relevance." In *Languages for Specific Purposes in History*, edited by Nolwena Monnier, 1–13. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Van Obberghen, Mathieu, Rita Temmerman, and Koen Kerremans. 2018. "Simultaneous Interpretation of Neosemanticisms in EU Press Conferences: Translations of 'Hotspot', 'Relocation' and 'Resettlement' into Dutch and German." *CLINA. An Interdisciplinary Journal of Translation, Interpreting and Intercultural Communication* 4 (1): 59–81.
- Vasiljevs, Andrejs, and Signe Rirdance. 2008. "Application of Terminology Standards for a Multilingual Term Bank: the EuroTermBank Experience." *Proceedings of the LREC–2008 Workshop on Uses and Usage of Language Resource-related Standards*, edited by Andreas Witt, Felix Sasaki, Elke Teich, Nicoletta Calzolari, and Peter Wittenburg, 55–60. Luxembourg: European Language Resources Association.
- Vecchi, Dardo de. 2016. "Approche pragmatérminologique des termes des entreprises et des organisations." *Synergies Italie* 12: 125–139.
- Villazón-Terrazas, Boris, Mari Carmen Suárez-Figueroa, and Asunción Gómez-Pérez. 2010. "A Pattern-Based Method for Re-Engineering Non-Ontological Resources into Ontologies." In *International Journal on Semantic Web and Information Systems* 6 (4): 27–63. <https://doi.org/10.4018/jswis.2010100102>
- Vinay, Jean-Paul, and Jean Darbelnet. 1995. *Comparative Stylistics of French and English: A Methodology for Translation*, translated by Juan C. Sager, and Marie J. Hamel. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/btl.11>
- Vossen, Piek. (ed.). 2002. *EuroWordNet: General Document*. University of Amsterdam (<https://research.vu.nl/en/publications/eurowordnet-general-document>). Accessed 8 December 2021.

W

- Warnier, Maxime. 2018. "Contribution de la linguistique de corpus à la constitution de langues contrôlées pour la rédaction technique : l'exemple des exigences de projets spatiaux." PhD diss., Université Toulouse Jean Jaurès.
- Warnier, Maxime, and Anne Condamines. 2015. "A Methodology for Identifying Terms and Patterns Specific to Requirements as a Textual Genre Using Automated Tools." In *Proceedings of the 11th International Conference on Terminology and Artificial Intelligence*, edited by Thierry Poibeau, and Pamela Faber, 183–190. Granada, Spain.
- Wauquier, Marine, Cécile Fabre, and Nabil Hathout. 2020. "Différenciation des noms d'action dérivés: le facteur de technicité étudié en corpus." In *Les corpus numériques à la modélisation linguistique en langues de spécialité*, edited by Cécile Frérot, and Mojca Pecman, 1–13. Grenoble: UGA éditions.
- Weilgaard Christensen, Lotte. 2019. "Danish Knowledge Patterns and Word Sketches for Semi-automatic Extraction of Terminological Information." In *New Challenges for Research on Language for Special Purposes: Selected Proceedings from the 21st LSP Conference*, edited by Ingrid Simonæs, Øivin Andersen, and Klaus Schubert, 173–189. Berlin: Frank & Timme.

- Weisgerber, Leo. 1958. "Ein Markstein angewandter Sprachwissenschaft: Begegnung mit Eugen Wüster." *Sprachforum* 3 (2): 92–95.
- Weissenhofer, Peter. 1994. "Zur Rolle der terminologischen Begriffslehre in der Translationswissenschaft." In *Translation Studies: An Interdiscipline*, edited by Mary Snell-Hornby, Franz Pöchhacker, and Klaus Kaindl, 319–326. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/btl.2.38wei>
- Werlich, Egon. 1975. *Typologie der Texte*. Munich: Fink.
- Werlich, Egon. 1976. *A Text Grammar of English*. Heidelberg: Quelle & Meyer.
- White-Hayes, Michael. 1996. *The Use of Metaphor in the British Press Reporting of the 1992 Currency Crisis*. Madrid: Universidad Complutense de Madrid.
- Whorf, Benjamin Lee. 1956. *Language, Thought, and Reality. Selected Writings of Benjamin Lee Whorf*, edited by John Carroll. Cambridge, MA: MIT Press.
- Wikipedia. "Collective Memory." (https://en.wikipedia.org/wiki/Collective_memory). Accessed 8 December 2021.
- Wilks, Yorick, Brian M. Sator, and Louise Guthrie. 1996. *Electronic Words: Dictionaries, Computers, and Meanings*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/2663.001.0001>
- Williams, Geoffrey (ed.). 2005. *La linguistique de corpus*. Rennes: Presses Universitaires de Rennes.
- Winston, Morton E., Roger Chaffin, and Douglas Herrmann. 1987. "A Taxonomy of Part-Whole Relations." *Cognitive Science* 11 (4): 417–444. https://doi.org/10.1207/s15516709cog1104_2
- Wisniewski, Edward J. 2002. "Concepts and Categorization." in *Steven's Handbook of Experimental Psychology: Memory and Cognitive Processes*, edited by Hal Pashler, and Douglas Medin, 467–531. Hoboken, NJ: John Wiley & Sons Inc. <https://doi.org/10.1002/0471214426.pas0211>
- Woyde, Rick. 2001. "Introduction to the SAE J2450 Translation Quality Metric." *Language International* April 2001: 37–39.
- Wright, Sue Ellen. 1997. "Representation of Concept Systems." In *Handbook of Terminology Management, Vol. I*, edited by Sue Ellen Wright, and Gerhard Budin, 89–97. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/z.htm1.13wri>
- Wright, Sue Ellen, and Gerhard Budin. 1994. "Data Elements in Terminological Entries: An Empirical Study." *Terminology* 1 (1): 41–59. <https://doi.org/10.1075/term.1.1.05wri>
- Wüster, Eugen. 1963 [1959]. "Die Struktur der sprachlichen Begriffswelt und ihre Darstellung in Wörterbüchern." In *Proceedings of the IIIrd Congress of the International Federation of Translators*, 615–627. Oxford/London/New York/Paris, Pergamon Press.
- Wüster, Eugen. 1968. *The Machine Tool: An Interlingual Dictionary of Basic Concepts Comprising an Alphabetical Dictionary and a Classified Vocabulary with Definitions and Illustrations. English-French Master Volume*. London: Technical Press.
- Wüster, Eugen. 1970 [1931]. *Internationale Sprachnormung in der Technik, besonders in der Elektrotechnik: die nationale Sprachnormung und ihre Verallgemeinerung*. 3rd ed. Bonn: Bouvier.
- Wüster, Eugen. 1971a. "Grundsätze der fachsprachlichen Normung." *Muttersprache* 5 (81): 289–295.
- Wüster, Eugen. 1971b. "Begriffs- und Themaklassifikationen. [Concept and Theme Classifications]." *Nachrichten für Dokumentation* 22(3): 98–104.
- Wüster, Eugen. 1973. Letter to Wolfgang Dressler, Typescript.
- Wüster, Eugen. 1974. "The Road to Infoterm: Two Reports Prepared on Behalf of UNESCO." Munich: Verlag Dokumentation.

- Wüster, Eugen. 1975. "Die Ausbildung in Terminologie und terminologischer Lexikographie." *Lebende Sprachen* 1975 (2): 33–38. <https://doi.org/10.1515/les.1975.20.2.33>
- Wüster, Eugen. 1976. "La théorie générale de la terminologie – un domaine interdisciplinaire impliquant la linguistique, la logique, l'ontologie, l'informatique et les sciences des objets." In *Essai de définition de la terminologie. Actes du colloque international de terminologie*, edited by Henriette Dupuis, 49–57. Québec: Régie de la Langue Française.
- Wüster, Eugen. 1979. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. Vienna/New York: Springer.
- Wüster, Eugen. 1985 [1979]. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*, 2nd ed. The LSP Centre, UNESCO Alsed Lsp Network, The Copenhagen School of Economics, Copenhagen.
- Wüster, Eugen. 1991 [1979]. *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. 3rd ed. Bonn: Romanistischer Verlag.
- Wüster, Eugen. 1993 [1974]. "Die Allgemeine Terminologielehre – ein Grenzgebiet zwischen Sprachwissenschaft, Logik, Ontologie, Informatik und den Sachwissenschaften." *Linguistics* 12 (119): 61–106. Reprinted in Heribert Picht, and Christer Laurén (eds). 1993. *Ausgewählte Texte zur Terminologie*, 331–376. Vienna: Termnet.
- Wüster, Eugen. 2001 [1959–1960]. "Das Worten der Welt, schaubildlich und terminologisch dargestellt. Leo Weisgerber zum 60. Geburtstag." *Sprachforum* 3/4: 183–204. Reprinted in Heribert Picht. 2001. *Terminologie und Wissensordnung –. Ausgewählte Schriften aus dem Gesamtwerk von Eugen Wüster*, 21–51. Vienna: TermNet.

Y

- Yong, Heming, and Jing Peng. 2007. *Bilingual Lexicography from a Communicative Perspective*. Amsterdam / Philadelphia: John Benjamins. <https://doi.org/10.1075/tlrp.9>
- Yu, Ning. 2008. "The Relationship between Metaphor, Body and Culture." In *Body, Language and Mind (Vol. II): Sociocultural Situatedness*, edited by Dirk Geeraerts, René Dirven, and John Taylor, 363–386. Berlin/New York: Mouton de Gruyter.
- Yu, Yating, and Dennis Tay. 2020. "A Mixed-Method Analysis of Image-Schematic Metaphors in Describing Anger, Anxiety and Depression." *Metaphor and the Social World* 10 (2): 253–272. <https://doi.org/10.1075/msw.00006.yu>

Z

- Zanola, Maria Teresa. 2014. *Arts et métiers au XVIIIème siècle, études de terminologie diachronique*. Paris: L'Harmattan.
- Zarnikhi, Abolfazl. 2014. "Towards a Model for Terminology Planning." PhD diss., University of Pompeu Fabra.
- Zawada, Britta E., and Piet H. Swanepoel. 1994. "On the Empirical Inadequacy of Terminological Concept Theories: A Case for Prototype Theory." *Terminology* 1 (2): 253–276. <https://doi.org/10.1075/term.1.2.03zaw>
- Zelinsky-Wibbelt, Cornelia. 2012. "Identifying Term Candidates through Adjective-Noun Constructions in English." *Terminology* 18 (2): 226–243. <https://doi.org/10.1075/term.18.2.04zel>
- Zetsche, Jost. 2008. "Geek Speak: A Maze of TEnTs." *ATA Chronicle* July 2008: 47 (https://www.ata-chronicle.online/wp-content/uploads/3707_47_geekspeak_column.pdf). Accessed 8 December 2021.

- Zgusta, Ladislav. 1971. *Manual of Lexicography*. Prague/The Hague: Academia/ Mouton. <https://doi.org/10.1515/9783111349183>
- Zhang, Weiwei, Dirk Speelman, and Dirk Geeraerts. 2015. "Cross-Linguistic Variation in Metonymies for PERSON: A Chinese-English Contrastive Study." *Review of Cognitive Linguistics* 13 (1): 220–256. <https://doi.org/10.1075/rcl.13.1.09zha>
- Zhu, Jianhua. 2007. "Terminologieforschung von Eugen Wüster – Frühblüher der Fachsprachenlinguistik." In *Universalitätsanspruch und partikulare Wirklichkeiten*, edited by Naoji Kimura, and Karin Moser V. Filseck, 251–255. Würzburg: Königshausen & Neumann.
- Zinken, Jörg, Lina Hellsten, and Brigitte Nerlich. 2003. "What is Cultural about Conceptual Metaphors?" *International Journal of Communication* 13 (1–2): 5–29.
- Zlatev, Jordan, Johan Blomberg, and Ulf Magnusson. 2012. "Metaphors and Subjective Experience: Motion-Emotion Metaphors in English, Swedish, Bulgarian and Thai." In *Moving Ourselves, Moving Others: Motion and Emotion in Consciousness, Intersubjectivity and Language*, edited by Ad Foolen, Ulrike Lüdtke, Timothy Racine, and Jordan Zlatev, 423–450. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/ceb.6.17zla>
- Zlatev, Jordan, Göran Jacobsson, and Liina Paju. 2021. "Desiderata for Metaphor Theory, the Motivation and Sedimentation Model and Motion-Emotion Metaphores." In *Figurative Language – Intersubjectivity and Usage*, edited by Augusto Soares da Silva, 41–74. Amsterdam/Philadelphia: John Benjamins. <https://doi.org/10.1075/ftl.11.02zla>
- Zorrilla-Agut, Paula, and Thierry Fontenelle. 2019. "Modernising the EU's IATE Terminological Database to Respond to the Challenges of Today's Translation World and beyond." *Terminology* 25 (2): 146–174. <https://doi.org/10.1075/term.00034.zor>

Standards and resources cited in chapters

A

ATA (American Translators Association). 2017. "ATA Certification Program Framework for Standardized Error Marking." (https://www.atanet.org/certification/Framework_2017.pdf). Accessed 23 December 2020.

ASTM (ASTM Intl., formally American Society for Testing and Materials)

ASTM. *ASTM DEFONLINE*: "ASTM Online Terminology Dictionary of Engineering Science and Technology." (<https://www.astm.org/bookstore/comps/defonline.htm>). Accessed 23 December 2020.

ASTM WK46396 (under development). ASTM Intl. *Analytic Evaluation of Translation Quality*. West Conshohocken, PA: ASTM.

ASTM WK54884 (under development). ASTM Intl. *Holistic Quality Evaluation System for Translation*. West Conshohocken, PA: ASTM.

B

Bio2RDF. 2020. (<https://bio2rdf.org/>). Accessed 16 June 2020.

BioPortal. 2020. (<https://biportal.bioontology.org/>). Accessed 16 June 2020.

C

CAS. (formerly Chemical Abstract Services). "CAS Registry – The Gold Standard for Chemical Substance Information." (<https://www.cas.org/support/documentation/chemical-substances>). Accessed 27 January 2021.

CLARIN (Common Language Resources and Technology Infrastructure). 2015-present. "CLARIN Concept Registry Browser." (<https://concepts.clarin.eu/ccr/browser/>). Accessed 5 January 2020.

Coreon¹. (<https://www.coreon.com/#:~:text=Coreon%20combines%20taxonomies%20with%20terminology,IoT%20applications%20interoperable>). Accessed 15 January 2021.

COTSOES (Conference of Translation Services of European States). 2003. *Recommendations for Terminology Work*. Bern: MediaCenter of the Confederation. (http://www.cotsoes.org/sites/default/files/CST_Recommendations_for_Terminology_Work.pdf). Accessed 26 February 2021.

D

DatCatInfo. Data Category Repository (DCR). Accessed January 16, 2020. <http://datcatinfo.net>

DFKI. 2015. "Multidimensional Quality Metrics (MQM) Issue Types." *Deutsches Forschungszentrum für Künstliche Intelligenz GmbH / German Research Center for Artificial Intelligence (DFKI) and QTLaunchPad* (<http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html>). Accessed 26 February 2021.

DiCoEnviro, *Dictionnaire fondamental de l'environnement*. 2020. (<http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search.cgi>). Accessed 30 October 2020.

DiCoInfo, *Dictionnaire fondamental de l'informatique et de l'Internet*. 2020. (<http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi>). Accessed 30 October.

DIN (Deutsches Institut für Normung)

DIN 2330–1979. Deutsches Institut für Normung. 1979. *Begriffe und Benennungen. Allgemeine Grundsätze*. Deutsche Normen. Berlin: Deutsches Institut für Normung e.V.

DIN 2331–1980. Deutsches Institut für Normung. 1980. *Begriffssysteme und ihre Darstellung*. Deutsche Normen. Berlin: Deutsches Institut für Normung e.V.

Dragon1. 2021. “Single Source of Truth.” *Concepts*. (<https://www.dragon1.com/concepts/single-source-of-truth>). Accessed 26 February 2021.

E

EC Innovations. “The Translation Quality Metric Widely Used by Automotive Manufacturers.” (<https://www.ecinnovations.com/blog/j2450-the-translation-quality-metric-widely-used-by-leading-auto-manufacturers/>). Accessed 11 January 2019.

eCl@ss. “The ECLASS Standard.” Accessed January 4, 2020. (<https://www.eclass.eu/en/standard.html>). Accessed 26 February 2021.

EcoLexicon. 2019. (<http://ecolexicon.ugr.es>). Accessed 26 February 2021.

EUR-Lex (European Parliament, Council of the European Union). 2014. Regulation (EU) No 516/2014 of the European Parliament and of the Council of the 16 April 2014 Establishing the Asylum, Migration and Integration Fund, amending Council Decision 2008/381/EC and Repealing Decisions No 573/2007/EC and No 575/2007/EC of the European Parliament and of the Council and Council Decision 2007/435/EC. EUR-Lex No. 516/2014. *Official Journal of the European Union*.

Euro TermBank Consortium. 2006. *Towards Consolidation of European Terminology Resources: Experience and Recommendations from EuroTermBank Project*. Riga: Tilde (https://www.eurotermbank.com/downloads/EuroTermBank_Towards_Consolidation_of_European_Terminology_Resources.pdf). Accessed 26 February 2021.

F

FIT (Fédération Internationale des Traducteurs). 2012. “Online Browsing Platform Published.” (<https://www.fit-ift.org/online-browsing-platform/>). Accessed 26 November 2020.

FrameNet. 2020. (<https://framenet.icsi.berkeley.edu/fndrupal/>). Accessed 20 October 2020.

G

GEMET (General Multilingual Environmental Thesaurus). 2020. (<https://www.eionet.europa.eu/gemet/en/themes/>). Accessed 20 October 2020.

GenSim. 2020. (<https://pypi.org/project/gensim/>). Accessed 26 February 2021.

I

IATE (Inter-Active Terminology for Europe). 2020. (<https://iate.europa.eu/home>). Accessed 15 February 2021.

IEC (International Electrotechnical Commission). 2020. *Electropedia: The World's Online Electrotechnical Vocabulary*. (<http://www.electropedia.org/>). Accessed 21 March 2020.

- IEC (International Electrotechnical Commission). "Common Data Dictionary (CDD)." (<https://cdd.iec.ch/cdd/iec61360/iec61360.nsf/TreeFrameset?OpenFrameSet>). Accessed 26 November 2020.
- INCa. 2013. "Cancer Info." *Institut National du Cancer* (<http://www.e-cancer.fr/Patients-et-proches/Cancer-Info>). Accessed 26 February 2021.
- InfoTerm. 2009. "New tool for standards development and use of standards: The ISO Concept Database" (ISO/CDB)." (http://www.infoterm.info/pdf/activities/news/ISOConceptDatabaseFlyer_en.pdf). Accessed 26 February 2021.
- interverbumTech. "TermWeb." (<https://interverbumtech.com/products-services/termweb/>). Accessed 23 December 2020.
- International Terminology Working Group. 1997. Guidelines for Forming Language Equivalents: A Model Based on the Art and Architecture Thesaurus. (<http://glotta.ntua.gr/StateoftheArt/Multilingual-Image-Retrieval/AAT/index.html>). Accessed 26 February 2021.
- ISO (International Organization for Standardization)**
- ISO 639. International Organization for Standardization (ISO). 1998–2021. *Codes for the Representation of Names of Languages*. Parts 1–4. Geneva: ISO.
- ISO 704:2009. International Organization for Standardization (ISO). 2009. *Terminology Work – Principles and Methods*. International Organization for Standardization. Geneva: ISO.
- ISO 860:2007. International Organization for Standardization (ISO). 2007. *Terminology Work – Harmonization of Concepts and Terms*. Geneva: ISO.
- ISO 1087:2019. International Organization for Standardization (ISO). 2019. *Terminology Work and Terminology Science – Vocabulary*. Geneva: ISO.
- ISO 1087–1:2000. International Organization for Standardization (ISO). 2000. *Terminology Work – Vocabulary – Part 1: Theory and Application*. Geneva: ISO.
- ISO 3166. International Organization for Standardization (ISO). *The International Standard for Country Codes and Codes for their Subdivisions*. Geneva, ISO.
- ISO 8879:1986. International Organization for Standardization (ISO). 1986. *Standard Generalized Markup Language*. Geneva: ISO.
- ISO 12200:1999. International Organization for Standardization (ISO). 1999. *Computer Applications in Terminology – Machine-readable Terminology Interchange Format (MARTIF) – Negotiated Interchange* (withdrawn). Geneva: ISO.
- ISO 10241–1:2011. International Organization for Standardization (ISO). 2011. *Terminological Entries in Standards – General Requirements and Examples of Presentation*. Geneva: ISO.
- ISO 10241–2:2012. International Organization for Standardization (ISO). 2012. *Terminological Entries in Standards – Part 2: Adoption of Standardized Terminological Entries*. Geneva: ISO.
- ISO/12616–1:2021. International Organization for Standardization (ISO). 2021. *Terminology Work in Support of Multilingual Communication – Part 1: Fundamentals of Translation-oriented Terminography*. Geneva: ISO.
- ISO 12620:2009. International Organization for Standardization (ISO). 2009. *Terminology and other Language and Content Resources – Specification of Data Categories and Management of a Data Category Registry for language resources*. Geneva: ISO.
- ISO 16642:2017. International Organization for Standardization (ISO). 2007. *Computer Applications in Terminology – Terminological Markup Framework*. Geneva: ISO.
- ISO 20539:2019. International Organization for Standardization (ISO). 2019. *Translation, Interpreting and Related Technology – Vocabulary*. Geneva: ISO.

- ISO 21720:2017. International Organization for Standardization (ISO). 2017. *XLIFF (XML Localisation Interchange File Format)*. Geneva: ISO.
- ISO 23185:2009. International Organization for Standardization (ISO). 2009. *Assessment and Benchmarking of Terminological Resources – General Concepts, Principles and Requirements*. Geneva: ISO.
- ISO 24156–1:2014. International Organization for Standardization (ISO). 2014. *Graphic Notations for Concept Modelling in Terminology Work and its Relationship with UML – Part 1: Guidelines for Using UML Notation in Terminology Work*. Geneva: ISO.
- ISO 24613–1:2019. International Organization for Standardization (ISO). 2019. *Language Resource Management – Lexical Markup Framework (LMF) – Part 1: Core model*.
- ISO 24619:2011. International Organization for Standardization (ISO). 2011. *Language Resource Management – Persistent Identification and Sustainable Access (PISA)*. Geneva: ISO.
- ISO 25964–1:2011. International Organization for Standardization (ISO). 2011. *Information and Documentation – Thesauri and Interoperability with Other Vocabularies – Part 1: Thesauri for Information Retrieval*. Geneva: ISO.
- ISO 26162–1:2019. International Organization for Standardization (ISO). 2019. *Management of Terminology – Terminology Databases – Part 1: Design*. Geneva: ISO.
- ISO 26162–2:2019. International Organization for Standardization (ISO). 2019. *Management of Terminology Resources – Terminology Databases – Part 2: Software*. Geneva: ISO.
- ISO/CD 26162–3 (under development). International Organization for Standardization (ISO). *Management of Terminology Resources – Terminology Databases – Part 3: Content*. International Organization for Standardization (ISO). Geneva: ISO.
- ISO 29383:2010. International Organization for Standardization (ISO). 2010. *Terminology Policies – Development and Implementation*. Geneva: ISO.
- ISO 30042:2019. International Organization for Standardization (ISO). 2019. *Management of Terminology Resources – TermBase eXchange (TBX)*. Geneva: ISO.
- ISO/AWI TS 24634:2020. International Organization for Standardization (ISO). 2020. *Management of Terminology Resources – Representation of Concept Relations and Subject Fields in TBX*. Geneva: ISO.
- ISO/CD 704. 2009. International Organization for Standardization (ISO). 2009. *Terminology Work – Principles and Methods*. Geneva: ISO.
- ISO/CD 704:2020. International Organization for Standardization (ISO). 2020. *Terminology Work – Principles and Methods*. Geneva: ISO.
- ISO/CD 12620–1:2020. International Organization for Standardization (ISO). 2020. *Management of Terminology Resources – Data categories – Part 1: Specifications*. Geneva: ISO.
- ISO/DIS 12620–2:2020. International Organization for Standardization (ISO). 2020. *Management of Terminology Resources – Data Categories – Part 2: Repositories*. Geneva: ISO.
- ISO/IEC/JTC 1/SC 32/WG 2, ISO/IEC AWI 5394. *Information Technology – Criteria for concept systems*.
- ISO/IEC/JTC 1/SC 32. ISO/IEC 11179-x. *Information Technology – Metadata Registries (MDR) – (Multipart standard)*. (<https://www.iso.org/committee/45342/x/catalogue/p/0/u/1/w/0/d/0>). Accessed 15 January 2021.
- ISO (International Organization for Standardization). “Online Browsing Platform (OBP).” (<https://www.iso.org/obp/ui/>). Accessed 26 November 2020.
- ISO (International Organization for Standardization). Smart Manufacturing Coordinating Committee (SMCC) (<https://www.iso.org/committee/54996.html>). Accessed 10 February 2021.

- ISO WD 24183:2021. International Organization for Standardization (ISO). 2021. *Technical Communication – Vocabulary*. Geneva: ISO.
- ISO/TR 22134: 2007. International Organization for Standardization (ISO). 2007. *Lignes directrices pratiques pour la socioterminologie*. Geneva: ISO.
- ISO TS 5060 (under development). International Organization for Standardization (ISO). *Translation Services – Assessment [Evaluation] of Translation Output – General Guidance*. Geneva: ISO.
- ISO/WD TR 24633:2020. International Organization for Standardization (ISO). 2020. *Building an RNG Schema for TBX core*. Geneva: ISO.

L

- Larousse*, s.v. “concept” (<https://www.larousse.fr/dictionnaires/francais/concept/17875#:~:text=Id%C3%A9e%20g%C3%A9n%C3%A9rale%20et%20abstraite%20que,d'en%20organiser%20les%20connaissances>). Accessed 17 December 2020.
- LISA. “LISA QA Model.” (<https://docs.sdl.com/785465/788071/sdl-multitrans/the-lisa-qa-model>). Accessed 10 January 2021.
- Litttré, Dictionnaire de français*, s.v. “souris” ([http://littre.reverso.net/dictionnaire-francais/definition/souris\[2\]/69580](http://littre.reverso.net/dictionnaire-francais/definition/souris[2]/69580)). Accessed 11 November 2020.

M

- Merriam-Webster. 2020. (<https://www.merriam-webster.com/>). Accessed 20 December 2020.

N

- NeoVisual*. 2020. (<http://olst.ling.umontreal.ca/dicoenviro/neovisual/>). Accessed 31 October 2020.

O

- OBO Foundry. 2020. *The Open Biological and Biomedical Ontology (OBO) Foundry*. (<http://www.obofoundry.org/>). Accessed 16 June 2020.
- O.H.I. 1994–1998. *Hydrographic Dictionary / Diccionario Hidrográfico / Dictionnaire hydrographique*. Monaco: Bureau hydrographique international.
- OMG (Object Modeling Group)
- OMG (Object Modeling Group). 2018. *DOL™, the Distributed Ontology, Model, and Specification Language™* (<https://www.omg.org/spec/DOL/1.0/PDF>). Accessed 26 February 2021.
- OMG* (Object Management Group). 2017. *Unified Modeling Language* (OMG UML*)*. (<https://www.omg.org/spec/UML/2.5.1/PDF>). Accessed 26 February 2021.
- OWL. 2012. *Web Ontology Language* (<https://www.w3.org/2001/sw/wiki/OWL>). Accessed 26 February 2021.

P

- POINTER. 2011. “Terminology Resources.” In *Pointer Final Report*, edited by Khurshid Ahmad, Robin Bon throne, Gert Engel, Aggeliki Fotopoulou, Deborah Fry, Christian Galinski, et al. University of Surrey. (<http://www.computing.surrey.ac.uk/ai/pointer/report/intro.html>). Accessed 26 February 2021.
- Prêt-à-LLOD (<https://pret-a-llood.github.io/>). Accessed 21 August 2020.

R

RDF. 2020. *RDF Schema 1.1*. (https://www.w3.org/TR/rdf-schema/#ch_label). Accessed 28 December 2020.

S

SAE (SAE International, formally Society of Automotive Engineers)

SAE J2450. 2001. *Translation Quality Metric*. Last revised August 17, 2016. Vehicle E/E System Diagnostic Standards Committee: Troy, Michigan.

SIL. "Ethnologue: Languages of the World." (<https://www.ethnologue.com/>). Accessed 24 January 2021.

SUMO. 2020. Suggested Upper Merged Ontology (<http://www.adampease.org/OP/>). Accessed 16 June 2020.

T

TAUS. June 15, 2015. "DQF and MQM Harmonized to Create an Industry-Wide Quality Standard." (<https://www.taus.net/academy/news/press-release/dqf-and-mqm-harmonized-to-create-an-industry-wide-quality-standard>). Accessed 26 February 2021.

Termcat (<https://www.termcat.cat/>). Accessed 21 August 2020.

TerminOrgs. "TBX Basic Specification." (<http://www.terminorgs.net/TBX-Basic.html>). Accessed 10 January 2021.

TERMIUM Plus®. 2020. (<https://www.btb.termiumplus.gc.ca/>). Accessed 12 December 2020.

TermWeb* (<https://interverbumtech.com/products-services/termweb/>). Accessed 10 January 2021.

TheMQM. 2021. *TheMQM Core Error Typology* (<https://themqm.info/typology/>). Accessed 1 November 2021.

U

Unesco (United Nations Educational, Scientific and Cultural Organization). 2017. "Integration." *Glossary* (<http://www.unesco.org/new/en/social-and-human-sciences/themes/international-migration/glossary/integration/#topPage>). Accessed 26 February 2021.

Unified Medical Language System (UMLS). 2020. (<https://www.nlm.nih.gov/research/umls/index.html>). Accessed 16 June 2020.

V

VerbNet. 2020. (<https://verbs.colorado.edu/verbnet/>). Accessed 29 October 2020.

W

Word2Vec. 2013. (<https://code.google.com/archive/p/word2vec/>). Accessed 26 February 2021.

W3C (World Wide Web Consortium)

W3C (<https://www.w3.org/>). Accessed 21 August 2020.

W3C. 2020. *Data Catalog Vocabulary (DCAT) – Version 2; W3C Recommendation* (<https://www.w3.org/TR/vocab-dcat-2/>). Accessed 10 February 2021.

W3C. 2009. *SKOS Simple Knowledge Organization System Reference* (<https://www.w3.org/TR/skos-reference/#L1045>). Accessed 26 February 2021.

- W3C-LEMON (Ontology Lexicon Community Group). 2016. Lexicon Model for Ontologies: Community Report." (<https://www.w3.org/2016/05/ontolex/>). Accessed 1 February 2021.
- W3C/MQM. 2020. "MQM Top Level (October 2020)." (<https://www.w3.org/community/mqmcg/mqm-top-level-october-2020/>). Accessed 26 February 2021.
- W3Cschools. "Introduction to XML." (https://www.w3schools.com/xml/xml_what_is.asp). Accessed 26 January 2020.
- W3C (OWL Working Group). 2012. *OWL 2 Web Ontology Language: Document Overview* (Second Edition) (<https://www.w3.org/TR/2012/REC-owl2-overview-20121211/>). Accessed 26 January 2020.
- W3Cschools. "XML RDF" (https://www.w3schools.com/xml/xml_rdf.asp). Accessed 28 January 2020.
- Wikilengua: Terminesp (<http://www.wikilengua.org/index.php/Wikilengua:Terminesp>). Accessed 21 August 2020.
- Wikipedia, s.v. "Pollutant." (<https://en.wikipedia.org/wiki/Pollutant>). Accessed 6 October 2020.
- Wikipedia, s.v. "System Integration." (https://en.wikipedia.org/wiki/System_integration). Accessed 26 February 2021.

Biographical notes

Ann Bertels is Associate Professor at KU Leuven (Belgium), where she teaches French for Specific Purposes at Leuven Language Institute (ILT). She coordinates ILT Teaching and is committed to the professionalisation of education. Her teaching and research interests include Quantitative Semantics and Quantitative Lexicology, Terminology and Language for Specific Purposes, specialised corpora, Pedagogical Lexicography and Business Communication. She has published numerous articles and conference contributions on Quantitative Semantics and Distributional Semantics in specialised corpora. The combination of Terminology, Semantics and Corpus Linguistics remains her main research interest. She is supervisor of doctoral research in the field of term variation and medical terminology. She also supervises master's theses with a focus on Business Communication.

Lynne Bowker holds a BA and MA in Translation from the University of Ottawa in Canada and a PhD in Language Engineering from the University of Manchester Institute of Science and Technology (UMIST) in the United Kingdom. Both her MA and PhD theses investigated aspects of multidimensional classification in Terminology, and she has continued to both teach and publish in the areas of Language for Special Purposes and Terminology Management. She is co-author (with Jennifer Pearson) of *Working with Specialized Language: A Practical Guide to Using Corpora* (2002, Routledge) and author of *Computer-Aided Translation Technology* (2002, University of Ottawa Press), which includes sections on terminology tools. She is also the editor of *Lexicography, Terminology and Translation: Text-based Studies in Honour of Ingrid Meyer* (2006, University of Ottawa Press).

Danielle Candel's main research areas belong to prescriptive Linguistics and Terminology and to Descriptive Lexicography. She has experience both in Lexicography (through her active participation in the writing of the CNRS dictionary *Trésor de la langue française*) and in Terminology (through extensive linguistic expertise and support for French Terminology and Neology Committees). Her publications focus on scientific and technical vocabularies, as well as on lexicographical and terminological theories and practices, specifically in the French institutional framework. After having been a research scientist in the CNRS laboratory "History of linguistic theories", University of Paris (Paris Diderot – Paris 7), she is now an associate member of this lab.

Anne Condamines is a senior researcher at the CNRS, in the CLLE (Cognition, Langues, Langage, Ergonomie) Laboratory. She is interested in the links between texts and terms and contributed to the definition of Textual Terminology in the early 1990s. At the same time, she co-created the “Terminology and Artificial Intelligence” research group. She continues to be interested in the possible synergies between these two disciplines. Her work has focused on the study of different types of terminological variation (in terms or phraseology) that can be identified in corpora: according to chronology, according to the degree of expertise, and according to disciplines. More recently, she has been interested in the construction of controlled languages using methods close to those of Textual Terminology. She has worked with various aeronautical and space organizations (Airbus, CNES (Centre National d'Etudes Spatiales), ENAC (Ecole Nationale de l'Aviation Civile).

Valérie Delavigne is a lecturer in Language Sciences at Sorbonne Nouvelle University, in the Department of Applied Foreign Languages, and a member of the Laboratory Clesthia (EA 7345). For several years, she has accompanied the editors of the information platform of the National Institute of Cancer (INCa) in France, where she directed the online dictionary for patients. Concerned with Linguistics in touch with reality, her research focuses on the use and social circulation of terminologies and on the forms of scientific, technical and medical popularization, in line with research on Socioterminology.

Marcel Diki-Kidiri was born in the Central African Republic, Marcel Diki-Kidiri (Ph.D.) studied primary and secondary education in the Central African Republic, Chad, the Republic of Congo, and Cameroon before pursuing university studies in Linguistics and Sociology of Language in France. He is a university Professor, a former Senior Researcher at the French National Center for Scientific Research, a high-level linguist specializing in Sanngo language, standardization and computer processing of languages, technical and scientific vocabularies. He initiated the Cultural Approach to Terminology laying the foundations for one of the current trends in the discipline. Academician of the African Academy of Languages (ACALAN) and Medalist of the Academic Palms of the Central African Republic, he also holds a CNRS Labor Medal (France).

Pascaline Dury is a professor at the Université Lumière Lyon 2. She teaches Terminology and Specialised Translation in the Department of Applied Foreign Languages. Her research focuses on the study of Terminology in specialised languages, particularly from a diachronic perspective. In this context, she is interested in the medical and textile fields. Her most recent work focuses on the terminology of emergence and innovation in these fields.

Pamela Faber is professor emeritus in the Department of Translation and Interpreting at the University of Granada, where she lectures on Terminology and Specialized Translation. She holds degrees from the University of North Carolina at Chapel Hill, the University of Paris IV, and the University of Granada. She is the director of the LexiCon research group, with whom she has carried out various national and international research projects that focus on applications of Frame-based Terminology to knowledge representation and the design of specialized knowledge bases. She is the author of over a hundred articles, chapters, and books on Terminology, Specialized Lexicography, and Lexical Semantics.

Sabela Fernandez-Silva holds a BA in Translation and Interpreting from the University of Salamanca, and a PhD in Applied Linguistics from Universitat Pompeu Fabra, Barcelona, Spain. She is associate professor at the Pontificia Universidad Católica de Valparaíso, Chile, and director of the Master's program in Applied Linguistics. She teaches courses on Terminology, Lexicology, Lexicography and linguistic theories to undergraduate and graduate students. Her research interests cover Terminology, Neology, Specialized Translation and Lexical Semantics from a cognitive-functional perspective. She is coordinator of the Research Group on Chilean Spanish Neology, which pertains to the research network "Antenas Neológicas". Her last research project was about the development of terminological knowledge in apprentices of a discipline.

Judit Freixa received her PhD in Philology at Universitat de Barcelona in 2002, with a thesis on terminology variation. Since 1994, she has been Full Professor of Terminology and Neology at the Universitat Pompeu Fabra (Barcelona), the head of the Neology Observatory, and a researcher at the Institute for Applied Linguistics (IULA) at the same university.

François Gaudin is professor of Linguistics at the University of Rouen with a PhD in History. As a member of the Laboratory "Langues, textes, discours, dictionnaires", he has developed Socioterminology in numerous articles and books, such as *Socioterminologie. Une approche sociolinguistique de la terminologie* (Duculot De Boeck, 2003). He has written biographies of Maurice Lachâtre (Lambert-Lucas, 2014) and Michel Morphy (Champion, 2021), and has conducted studies on the cultural history of dictionaries. His publications include *La lexicographie militante. Dictionnaires du XVIIIe au XXe siècle*, (Champion, 2013) and *Dictionnaires en procès* (Lambert-Lucas, 2015). He has also edited unpublished letters of Karl Marx and the Communards (2003).

John Humbley is an emeritus professor at the Université de Paris. His background is in Lexicology, as he worked for many years with Bernard Quemada, the pioneer of Applied Linguistics in France and major editor of the *Trésor de la langue française*. He has been active in Terminology since his involvement in the Centre de terminologie et de néologie as from 1987, in particular in connection with Infoterm (of which he was vice-president in the 1990's) and with the Réseau international de néologie et de terminologie. Neology has also been a major centre of interest, culminating in the founding of the journal *Neologica* in 2006, which he edited with Jean-François Sablayrolles until 2018. He is on the board or scientific committee of several journals such as *Terminology*, *ASp*, *Meta*, *Fachsprache*, *Les Cahiers de lexicologie* and publishes on Terminology, Neology and Specialized Translation.

Kyo Kageura, PhD, is Professor of the Library and Information Science Laboratory at the Graduate School of Education, University of Tokyo. He has authored several books, including *The Quantitative Analysis of the Dynamics and Structure of Terminologies* (John Benjamins, 2012), and has published extensively in international journals and conferences in the fields of Information Studies, Terminology, Computational Linguistics and Translation Studies. He also runs the online translation training system Minna no Honyaku Jisshu (<http://mnh-tt.org/>). He serves as an editor of the journal *Terminology* and the research monograph series *Terminology and Lexicography Research and Practice*, both published by John Benjamins, and as a committee member of a number of international conferences in the fields of Terminology, Library and Information Studies, and Language Processing.

Natalie Kübler is a full professor in Linguistics and Translation Studies at the Université de Paris. In her early career, she worked on corpus-based contrastive analysis of specialised verbs, learner corpora for error-correction software and writing-aids in English for French-speakers. She then designed a Web-based environment for ESP corpus-driven learning, the WALL (Web-Assisted Language Learning) project. She was the promoter of the MeLLANGE project, an EU-funded project with ten partners, in which one of the first multilingual, multidirectional and multi-genres learner translator corpus was developed. Her current interests are the specialisation of neural machine translation, post-editing learner corpora, and semantic prosody in translation. She has been the director of the CLILLAC-ARP research team since 2014.

Marie-Claude L'Homme is a full professor at the Department of Linguistics and Translation, University of Montreal, and a member of the Observatoire de linguistique Sens-Texte (OLST), a research group she headed for almost 10 years. Her research topics are Terminology Theory and Lexical Semantics and Corpus Linguistic applied to Terminology. She supervises the development of multilingual

terminological resources (DiCoInfo, DiCoEnviro, and Frame DiCoEnviro). She is the author of a French textbook in Terminology (*La terminologie : principes et techniques*; the second edition was published in 2020) and an English book on the application of Lexical Semantics to Terminology (published by John Benjamins in 2020). She was co-editor of Terminology for almost 20 years (2000-2019). She was appointed fellow of the Royal Society of Canada in 2013.

Pilar León Araúz lectures and works in Terminology, Translation, Corpus Linguistics and Translation Technologies. She holds degrees from Northumbria University (UK), the Université de Provence (France) and the University of Granada, where she has been a lecturer in Translation and Interpreting since 2007. Her research focuses on knowledge extraction and representation for the development of terminological resources, using corpus-based techniques. She has published papers in international journals such as *Terminology*, *International Journal of Lexicography* and *Frontiers* and book chapters in publishers such as Oxford, Routledge, Peter Lang, Springer or John Benjamins. She serves on the scientific boards of international journals and conferences.

Elizabeth Marshman holds a PhD in Translation (Terminology option) from the Université de Montréal. She is Associate Professor at the University of Ottawa's School of Translation and Interpretation, where she teaches primarily in areas relating to Terminology and Translation technologies. Her main research interests include corpus-based and computer-assisted terminology research, with a particular focus on the use of knowledge patterns in terminology research and description. Her other interests include user-centred evaluation of terminology resources and translation technologies, and the place of technologies in translator education.

Ingrid Meyer earned an MA in Translation and a PhD in Linguistics from the Université de Montréal and worked as a professor at the School of Translation and Interpretation at the University of Ottawa for over twenty years until the time of her death in 2004. A pioneer in knowledge-based and corpus-based terminology, Professor Meyer's design and development of the COGNITERM terminological knowledge base was funded by successive grants from the Social Sciences and Humanities Research Council of Canada. She published regularly on topics related to language for special purposes, including dictionary use, concept analysis, de-terminologization, metaphors in terminology, knowledge patterns and knowledge rich contexts, with her work appearing in journals such as *Terminology*, *International Journal of Corpus Linguistics*, and *Meta*, among others. An active member of the EURALEX association, she was a regular presenter at both the EURALEX and Terminology and Knowledge Engineering (TKE) conferences.

Elena Montiel-Ponsoda is Associate Professor of Applied Linguistics at the Universidad Politécnica de Madrid (UPM), in Madrid, Spain, and member of the Ontology Engineering Group at the same University. Her main research interests are in the common ground between Terminology and Ontology Engineering. Her main research has focused on the development of models to enrich ontologies with multilingual information and to explain terminologies and other language resources as linked data on the Web. She is currently exploring the use of inter-linked multilingual terminologies in semantic-based information retrieval. She has recently successfully coordinated the innovation action Lynx (a project funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No 780602).

Anita Nuopponen is Professor Emerita of Technical Communication in the School of Marketing and Communication, University of Vaasa, Finland. She was the head of Department of Communication Studies for fifteen years and has been teaching and developing Terminology courses for four decades. Her research interests include Terminology theory and practice, systematic concept analysis, usability and online communication. She has been heading a research group (Termface) that focuses on linguistic and terminological usability of user interfaces. She has written and co-authored over eighty articles and chapters on Terminology and Technical Communication, e.g. a Swedish textbook on the theory and practice of Terminology (*Ordning och reda – Terminologilära i teori och praktik* 2010, reprinted by Studentlitteratur 2016) together with professor Nina Pilke. Since 1994, she maintains the Terminology Forum website compiling links to Terminology-related information online.

Mojca Pecman is a full professor in Terminology and lexical resources creation at the Université de Paris. Her research focuses on terminology management and the analysis of language phenomena operating at the lexico-discursive level, which involve studies in a number of related disciplines: Terminology, Phraseology, Discourse Analysis, LSPs, Specialised Translation and dictionary-making. She is in charge of a research project on the creation of terminological and phraseological resources which led to the design of the ARTES online multilingual and multidomain database with a view to improving translation-related research and teaching. Her latest works focus on terminological variation in scientific discourse. She is the author of a book *Langue et construction de connaissances. Energie lexico-discursive et potentiel sémiotique des sciences* (L'Harmattan Editions, 2018), and since 2015 she serves on the advisory board of *Terminology* journal (John Benjamins publishing).

Aurélié Picton is an associate professor at the Faculty of Translation and Interpreting at the University of Geneva (Department of Translation Technology, referred to by its French acronym TIM) since 2011. She teaches courses in Terminology, Applied Lexicology, Languages for Special Purposes and corpora for translation. She is an associate member at the Observatoire de Linguistique Sens-Texte (OLST, Department of Linguistics and Translation, University of Montreal). Her main research interests cover two areas: the treatment of variation in specialized corpora and the ergonomics of translation, in particular the impact of technologies on language practitioners. She was a partner in the CRISTAL project (ANR 2012-2016) on the integration of corpus data in translation practice and is regularly involved in several applied projects (e.g. in collaboration with the Cnes and the University of Toulouse, the Geneva Centre in Humanitarian Studies (ex. CERAH), etc.).

Francisco Ruiz de Mendoza Ibáñez has headed seven research projects of the Spanish National R+ d+i Plan. He has supervised fifteen doctoral dissertations and published over 170 papers. He has (co-)edited eight books and authored or co-authored eight others, among them *Cognitive Modeling. A Linguistic Perspective* (with Alicia Galera) (2014, John Benjamins), which received the 2015 AESLA Senior Researchers' Award, *Ten Lectures on Cognitive Modeling. Between Grammar and Language-Based Inferencing* (2021, Brill), and *Modeling Irony* (with Inés Lozano) (2022fc, John Benjamins). He has been invited on over 200 occasions to lecture at scientific events across the world. Editor of the *Review of Cognitive Linguistics* (Benjamins), and co-editor of *Applications of Cognitive Linguistics* (Mouton), he has served on 35 advisory boards. He presided the *Spanish Association of Applied Linguistics* (2005-2011).

Rita Temmerman was full professor in Translation and Terminology and Multilingual Intercultural Communication at Erasmushogeschool Brussel and Vrije Universiteit Brussel until 2019 when she retired. Her research contributed to the Sociocognitive Approach in Terminology studies. She focuses on issues related to Special Language and Terminology, such as application-oriented multilingual terminology analysis and understanding terminology creation in Cognitive, Linguistic, cultural, and situational contexts. She has published on Cognitive Linguistics and Special Language, metaphor in cognition, dynamics and Terminology, intercultural and multilingual communication, secondary term formation, terminology harmonisation within the EU, sensory experience and descriptors, and translanguaging.

José Manuel Ureña Gómez-Moreno is an Associate Professor in the Department of English and German Philologies at the University of Granada (Spain), where he lectures in English Linguistics and Academic English. His main research interests lie in Cognitive Semantics, figurative language, Terminology and English-Spanish contrastive analysis. He is a member of the *LexiCon* research team, with which he has completed a good number of funded research projects. José Manuel is the author of a wide variety of articles published in high-impact academic journals, such as *Cognitive Linguistics*, *Review of Cognitive Linguistics*, *Terminology*, *Metaphor & Symbol*, *Language Sciences* and *Biosemiotics*. He is also the author of a number of book chapters published by leading international publishing companies, such as *John Benjamins* and *Mouton de Gruyter*.

Sue Ellen Wright Sue Ellen Wright is an Emerita Professor of German and former member of the Kent State University Department of Modern and Classical Language Studies. She taught Terminology Studies and computer applications for translators and German to English translation. She is a past chair of the American Translators Association (ATA) Terminology Committee and is ATA certified for German to English translation. She is engaged in national and international standards (ASTM and ISO) and chairs the U.S. mirror committee for ISO Technical Committee 37, *Language and Terminology*. Together with Professor Gerhard Budin of the University of Vienna she compiled the *Handbook for Terminology Management* and is the author of many articles on applied terminology management in industry. She has served as Chair of TC 37/SC3 and as Convenor for SC3/WG1 for Data Categories. She is a past recipient of the Eugen Wüster Prize and of the ATA Gode medal.

Index

2D-plot 314

3D-plot 314

A

accuracy 40, 52

activity relation 68, 78, 79

ad hoc terminology 92

adjective 240, 246, 249

AILA 38, 40

Allgemeine Terminologielehre
42, 46, 55

ambiguity 307

analytic quality evaluation 101

angle of referent perception
393

antonymy 253

Applied Linguistics 17, 18,
38, 194

appropriateness 462

appropriation 198, 199, 200,
201

archetype 198, 199, 201, 206,
210

argument 248, 249, 265

argument structure 240, 248,
250

Artificial Intelligence 32, 112,
220

associative relation, associative
relationship 70, 71, 92

automatic term extraction
464, 474

B

BabelNet 170

bag-of-word model 315

Begriffssystem 41, 45, 48

behaviour-based resemblance
389

Benennungssystem 42

BERT 318

BEST 140, 143, 144

Bidirectional Encoder

Representations from

Transformers 318

Bidirectional Encoder

Representations from

Transformers for Biomedical

Text Mining 318

bilingual terminology work 113

Bilingualized English-Spanish

Thesaurus 143

BioBERT 318

Bioinformatics 181

biunivocity 23, 48, 52, 66, 180,
220

borrowed constituent element
469

bottom-up analysis 225

BOW 315

C

CamemBERT 326

CamemBERT-bio 326

categorization 334, 338, 359,
435, 439

category 338

causal relation 70, 71, 83

cause-effect relation 117, 293

cause of terminological
variation 399, 400, 405

CBOW 318

CBTS 265

CCR 107

CDB 96

CDD 107

characteristic 64, 65, 73, 84,
128, 131

Chemistry 180

chronological variation 405

circulation of terms 184

CLARIN Concept Registry 107

class membership 133

class of objects 207

classical theory of classification
133

classification 95, 127, 130, 131,

137

classification scheme 19

classificatory pattern of term
formation 467

closed thematic terminology
113

clue, linguistic 228

CNL 230

CODE 123

COGNITERM 136, 158

cognition 335, 339, 350

cognitive approach to term
variation 438, 439

cognitive approach to terminology
25

cognitive cause of variation
405, 406, 412, 416

cognitive dimension of variation
437

cognitive factor of term variation
443, 444

cognitive function of term
variation 407, 443, 444

Cognitive Linguistics 379

cognitive motivation of term
variation 369, 441, 453

cognitive negotiation 183

collocate 247

collocation 247, 253

collocational strength 313

combination 464

Common Data Dictionary
107

Communicative and Textual
Terminology 16

communicative function of term
variation 449, 454

Communicative Theory of
Terminology 179, 198, 402

- community of practice 88
 - comparable corpora 229, 233
 - competency question 163
 - complex term 116
 - complex term formation 462, 463
 - composition 462
 - compositional sequence 251, 252
 - compositionality 253, 325
 - compound metaphor 382, 383, 388
 - compound relation 69, 73, 75
 - Computational Linguistics 150
 - Computer Science 22
 - concept 24, 28, 31, 37, 40, 41, 42, 48, 57, 128, 199, 204, 206, 207, 211, 212, 241, 332, 333, 399, 401, 407
 - Concept Database 96
 - concept flexibility 444, 445
 - concept formation 63, 65
 - concept management 113, 120
 - concept management technology 119
 - concept network 115
 - concept-oriented terminology management 88
 - concept relation, conceptual relation, conceptual relationship 64, 65, 66, 68, 69, 70, 71, 116, 131, 136, 295, 301
 - concept system 49, 64, 65, 81, 96, 129, 131, 134, 135, 480, 489, 490, 492
 - conceptual 112
 - conceptual analysis 114
 - conceptual characteristic 114, 117, 468
 - conceptual categorization 360, 435
 - conceptual category 469
 - conceptual conjunction 44
 - conceptual correspondence 490, 498
 - conceptual dimension 112, 120
 - conceptual disjunction 44
 - conceptual equivalence 487, 494
 - conceptual graph 145
 - conceptual incongruence 114, 496, 497
 - conceptual metaphor 379, 380, 385, 389
 - Conceptual Metaphor Theory 379, 384
 - conceptual modelling 353, 357, 358, 369
 - conceptual paradigm 155
 - conceptual specification pattern 468, 470
 - conceptual system 462
 - conceptual variation 401
 - conceptualisation, conceptualization 27, 403, 410, 412, 416, 418
 - Conceptually Oriented Descriptive Environment 123
 - continuous bag of words 318
 - concordancer 292
 - conjunction 44, 45, 51
 - consistency 462
 - constituent element 464
 - context 117
 - contextual definition 117
 - contiguity relation 68, 69, 73, 75
 - controlled natural language 229
 - conventional metaphor 380
 - co-occurrence analysis 311
 - co-occurrence matrix 313, 317
 - co-occurrence vector 313, 317
 - coordinate concepts 72, 117
 - CoP 88
 - corpus 27, 137, 140, 186
 - corpus analysis 191, 297
 - Corpus-based Terminology 138, 221, 264, 266, 268, 284, 311
 - Corpus-Based Translation Studies 265
 - corpus building 300
 - corpus design 274, 280
 - Corpus Linguistics 27, 187, 220, 221, 224, 264
 - corpus size 319, 324
 - correctness 52
 - correspondence 480, 488, 492, 493
 - cosine angle 314
 - count-based semantic model 316
 - creative metaphor 386, 387
 - cross-domain mapping 389
 - cultural profile 370, 374
 - Cultural Terminology 179, 197, 198, 200, 203, 212, 339
 - cultural variation 270
 - culture 198, 199, 200, 202, 203
 - culture-specific metaphor 393
 - culture-typical metaphor 393
- D**
- Data Catalog Vocabulary 107
 - data category 99
 - Data Category Repository 106
 - data modelling 95
 - data retrieval 120
 - data size 473
 - data sparseness 314, 324
 - DCR 106
 - deference 185
 - defining expositive 292
 - definition 117
 - definition construction 117, 125
 - degree of cultural specificity 393
 - degree of equivalence 478, 481, 483, 484, 487, 496
 - degree of precision 350
 - deliberate metaphor 383, 385, 387
 - denomination 44, 45, 47, 48, 51, 199, 204, 206, 207, 208, 211, 216
 - denominative variant 403, 404, 418
 - denominative variation 205, 400, 401, 403, 405, 407, 414
 - denotation 44
 - description 38
 - descriptive 38, 47
 - descriptive terminology 58
 - descriptivism 59
 - designation 92
 - designation paradigm 184
 - determination 44
 - determinologisation 224
 - developmental relation 70, 81
 - diachronic corpus 432, 433

- diachronic scale 472
 diachronic understanding 331
 diachronic variation 183, 405, 408, 410, 421, 423, 426, 429, 430
 diachrony 233
 dialectal cause of variation 405, 408
 diaphasic variation 406
 diastratic variation 405, 406, 408, 409
 diatopic variation 405, 406, 408
 DiCoEnviro 258
 dictionary 19, 20
 differentia 117
 dimension 112, 129, 130, 134, 135, 146
 dimension reduction 314, 317
 diplomacy 181
 directed graph 471
 discipline 264, 269, 280
 discourse 24, 34
 discourse analysis 181
 discourse community 264, 269
 discourse-oriented terminology management 87
 discursive cause of variation 405, 406
 disjunction 44
 distributional approach 225, 233
 Distributional Hypothesis 312
 distributional semantic analysis 312
 Distributional Semantic Model 313, 315
 Distributional Semantics 312
 distributional thesaurus 321
 division of linguistic labor 334, 350
 document-based semantic model 315
 document production chain 119
 documentation 19, 34, 114
 documentation selection 115
 domain 22, 183, 184, 264, 269, 278, 280, 301, 303
 domain expert 226, 227, 228
 domain knowledge 33, 292, 306
 domain-specific term 239
 domain-specific vocabulary expansion 312, 320, 326, 264, 266, 268
 DS 311
 DSM 313, 315
 dynamic structure 389
 dynamic understanding 335
 dynamics of categorization 437, 439
 dynamics of terminology 340
 E
 eCl@ss 108
 EcoLexicon 145, 159, 295, 320, 363, 372
 edge 471
 editology 186
 Eineindeutigkeit 48, 50, 52, 58
 Electropedia 155, 156, 162
 ELMo 318
 Embeddings from Language Models 318
 embodiment 19, 30, 263, 284, 334, 379, 439, 440
 emergent distributed understanding 335
 encyclopedic resource 88
 end-user 226, 227, 228
 entity 240
 environment 240
 Epistemology 18, 26
 equivalence 29, 478, 480, 481, 486, 487, 488, 490, 496, 499
 equivalent formation 480, 485, 498
 equivalent term 150, 497, 498
 Ergonomic Linguistics 227, 230
 error root cause 102
 Esperanto 20, 25, 38
 Ethnoterminology 179
 event 240
 exegetical metaphor 385
 experience 198, 200, 202
 experiential realism 439
 experientialism 379, 439
 expert 184
 expert knowledge 121
 expert opinion 122
 expert-to-expert communication 396
 Extended General Theory of Terminology 267
 extended semantic triangle 333
 eXtensible Markup Language 93
 extralinguistic elements 432, 433, 445
 F
 fachsprachliche Lexikologie 46
 Fachwörterbuch 43, 45
 FBT 141
 field 272, 279
 first-order co-occurrences 316
 flexibility 463
 flexible performance capability 343
 flexible terminological definition approach 146
 floating signifier 341, 346
 fluidity 125
 form of terms 468
 formal variant 451
 formality continuum 125
 forms of possible terms 464
 Founder of Terminology 16
 frame 246, 250, 323, 353, 354, 355, 356, 357
 Frame Semantics 26, 245, 323, 333, 356, 441
 Frame-based Terminology 134, 141, 354, 361
 framing 337
 FTDA 146
 functional cause of variation 405, 413
 fuzzy concept 125
 G
 General Terminology 40, 43, 46, 50, 54, 55
 General Theory of Terminology 16, 23, 37, 38, 50, 54, 56, 177, 188
 generic concept 65, 71
 generic relation, generic concept relation 69, 71, 72, 92
 generic-specific relation, generic-specific relationship 117, 130, 139, 141, 293, 299
 GENOMA-KB 159

- genre 190, 264, 267, 268, 269, 271, 272, 281
- genre-domain configuration 278, 279
- genre type 276, 278
- genus-differentia definition 117
- geographical fragmentation 395
- geographical variation 405
- Global Vector 317
- glottonomy 190
- Glottopolitics 178, 180
- GloVe 317
- good noise 306
- graphical context 372, 374
- graphical representation 137
- graphics 124
- grammatical knowledge pattern 292
- growth pattern of constituent elements 466, 470
- GTT 177
- H**
- HAL-model 322
- handbook 17, 18, 21, 22, 23
- healthcare 181
- heuristic metaphor 385
- hierarchical relation, hierarchical conceptual relation 69, 92, 117, 124, 136, 141, 144, 145
- hierarchy 135, 145
- high-dimensional matrix 317
- historicity 200, 203
- holistic quality evaluation 101
- human being 200, 202
- human-readable 87
- hypernomase 117, 494, 495, 499
- hyperonymy 294
- hypertext browsing 124
- hyponymy 130
- I**
- IATE 157, 158
- identity 180
- IEC 49
- image metaphor 381, 382, 389
- image-schematic metaphor 382, 388
- in-domain knowledge 325, 327
- indeterminacy 335
- individual 45
- information modelling 95
- Information Science 17
- inheritance 125, 135
- instrumental relation 68, 69, 78
- intensional definition 117
- interaction 179
- interactional perspective 180
- intercategorical information 451, 453, 454
- intercultural understanding 347
- interdisciplinary field 17
- interface discussion 183
- interlingual variation 396
- interlinguistic cause of variation 405
- interlinguistic equivalence 114, 128, 487, 498
- International Astronomical Union 185
- Internet of Things 100
- interpretation 222, 226, 228
- intertextuality 191
- intra-textual term variation 449
- intracategorical information 451, 453, 454
- intra lingual variation 396
- intralinguistic equivalence 477, 487, 498
- IoT 100
- is-a hierarchy 130
- ISA 40, 49
- ISO 38, 40, 49
- ISO SMART 107
- ISO TC 37 48
- ISOCat 106
- isolated term 472
- Ist-Norm 33, 42, 47, 52
- K**
- Keyword in Context 92
- knowledge 198, 199
- knowledge acquisition 121, 122
- knowledge base 27, 121, 122
- knowledge-based approach 238
- knowledge-based system 121
- knowledge-based terminology 129, 133, 135
- knowledge construction 283, 437, 449, 455
- Knowledge Engineering 112, 120, 121, 129, 219, 222, 231
- knowledge formalization 122
- knowledge modelling 95
- knowledge organization 88
- knowledge pattern 129, 139, 147, 292, 293, 295, 323, 361, 364, 374
- knowledge probe 292
- knowledge refinement 122
- knowledge resource 354, 355, 373, 374, 440
- knowledge-rich context 140, 165, 223, 226, 292, 364, 369, 374
- knowledge structure 115
- knowledge transfer 16, 66, 85, 449, 462
- KP 139
- KRC 292
- KWIC 92
- KWIC concordancer 297
- L**
- language contact 406, 415, 416
- language for special purposes 18, 39
- language planning 17, 23
- language policy 25
- language practice 183
- Latent Semantic Analysis 315, 317
- LBX 93
- legal integration 343
- level of specialisation 414
- Lexical Base eXchange 93
- lexical item 116, 459
- lexical knowledge pattern 292, 293
- Lexical Markup Framework 93, 167
- lexical pattern 139, 140
- lexical relation 241, 253
- Lexical Semantics 237
- lexical unit 241
- lexicographer 461
- lexicographical resource 88

- Lexicography 38, 42
lexico-semantic approach 238
lexicon 45, 243
linguistic 112
linguistic bridge 184
linguistic dimension 112
linguistic economy 462
linguistic insecurity 180
Linguistic Linked Open Data
94, 170
linguistic variation 41, 58
Linguistics 18, 25, 26, 28, 34,
237
link 132, 144
Linked Open Data 87
Linked Open Terms methodology
163
LLOD 94, 170, 171
LMF 93
loanword 405, 415
logical relation 68, 72
loosely coupled resource 91
low-dimensional semantic space
317
LSA 315
LSP 18, 39
- M**
machine learning 231, 233
machine parsable 88
machine-processable 87
machine-readable 94
Machine Readable Cataloguing
97
machine readable dictionary
93
machine-sensible 88
MARC 97
marine biology 384, 392
MDS 317
meaning 254
memory 200, 201, 203
meronymy 130, 294
metadata markup 103
metaphor 33
metaphor theory 29
metaphor type 385, 389, 395
metaphorical term 389
metaphorical understanding
335
microsense 243
misunderstanding 334
modelling terminological growth
466, 472
modification 464
monosemy 57, 180
morphological relation 249
MQM 101
MRD 93
multi-word expression 251, 253
multi-word sequence 240
multi-word term 247, 324, 325,
354, 366, 374, 375
multidimensional classification
132, 134
multidimensional knowledge
122
MultiDimensional Scaling 317
multidimensional space 314
multidimensionality 125, 128,
133, 136, 137, 138, 299, 307,
417, 445
multidisciplinarity 119
multilingual knowledge system
100
multilingual quality management
101
multilingual terminology work
118
multilingualism 120
Multimodal Distributional
Semantic Model 315
multiple inheritance 135
- N**
Named Entity Recognition 318
naming 438, 442
naming negotiation 183
native constituent element 469
Natural Language Processing
94, 219, 220
nature and nurture 391
nature of arguments 249
near synonymy 253, 255
necessary and sufficient
characteristics 134
negotiated understanding 335
neologism 118, 335
neology 423, 424, 428
NER 318
neural embeddings 317
neural network 318
neural network algorithm 317
new, the 198, 200, 201
new concept 462
NLP 94, 222
NMF 317
node 132, 139, 144, 471
noise 116, 305, 306
non-compositional sequence
252
non-equivalence 482, 483, 487,
490, 493
non-hierarchical relation 69,
124, 136, 141, 144, 145
non-resemblance metaphor
388, 389, 390
Nonnegative Matrix
Factorization 317
norm 16, 18, 28, 34, 268
normaison 186
notion 40
noun 242, 246
noun phrase 246
novel metaphor 386, 387
nuclear production 181
- O**
object 39, 40, 41
object-function relation 293
OBP 97
Oncology 181
OncoTerm 159
Online Browsing Platform 97
onomasiology 15
Ontolex-Lemon 94
ontological metaphor 381
ontological relation 70
ontology 31
ontology-based NLP
application 150, 160, 168
ontology building 312, 321
ontology implementation 163,
164, 165
ontology maintenance 163, 164
ontology publication 163, 164
ontology requirements
specification 163, 164
ontology reuse 165
OntoTerm 159
Ontoterminology 162
open thematic terminology 113
opposition 255

orientational metaphor 381, 382
 origination relation 70, 76
 Orthonymy 191
 OWL 103, 153, 154

P

paradigmatic relation 248, 249, 253
 paralinguistic knowledge
 pattern 292
 part-of hierarchy 130
 part-whole, part-whole relation
 74, 133, 140, 143, 146, 293, 295,
 300, 303, 305
 partial equivalence 483, 493, 494
 partitive relation 69, 71, 74, 92
 pattern set 297
 PCA 317
 percept 199, 205, 207, 212
 perception 199, 201, 204, 207
 Persistent Identifier 107
 Philosophy 19, 31
 phraseology 57, 247
 physical resemblance 389, 390
 PID 107
 polyphony 191
 polysemy 243, 257, 335
 popularizing 25
 Positive Pointwise Mutual
 Information 314
 Positivism 17, 20
 PPMI 314
 practical terminology work
 459
 Pragmaterminology 179
 precision 305, 306
 predicate-argument relation
 253
 predicate-argument structure
 367, 375
 predicative term 240
 predicative unit, predicative
 lexical unit 240, 248, 249,
 256
 prediction-based semantic
 model 316, 317
 prescription 59
 prescriptive 38, 47, 58
 prescriptive norm 48

Prescriptivism 59
 primacy of the concept 16, 30
 primary metaphor 380, 382,
 388, 390
 Primary Metaphor Theory
 380, 382
 primary term formation 462
 Principal Components Analysis
 317
 process of understanding 339
 property 240
 Prototype Theory 129, 133, 134,
 440, 441
 prototypical understanding
 338, 347
 pseudosynonym 118
 Psychoterminology 179
 PuertoTerm 140, 141

Q

qualitative approach to
 terminological growth 466
 quality control 118, 125
 quantitative approach to
 terminological growth
 466, 470
 quasi-predicative unit 240

R

Random Indexing 314, 321
 rank relation 69, 74, 76, 77
 RDF 94, 154, 158
 realistic possibility of existence
 465
 realistic range of possible terms
 466, 468
 recall 306
 reception theory 20
 reconceptualization 199, 202,
 205
 recontextualized network 146
 reduction 450, 451
 reference 185
 register 264, 268, 270, 271, 272
 relation marker 292, 293, 297
 relatively motivated 462
 representation of specialized
 knowledge 374, 437, 439,
 453
 research article 450

resemblance metaphor 382,
 388, 389, 390, 392
 Resource Description Framework
 94, 154
 revision 118
 RI 317
 role relation 68, 71

S

scale 213
 science popularization 183
 scientific article 279
 score card 102
 second-order co-occurrences
 316
 secondary term formation 462
 sector 202, 213
 semantic distinction 253
 semantic frame 323
 semantic network 355, 356,
 373, 489
 semantic ordering system 87
 semantic proximity 322
 semantic relatedness 311
 semantic relation 65, 68, 69,
 70, 291, 354, 358, 361, 362,
 363, 364
 semantic representation 355,
 356, 361
 semantic similarity 311
 semantic spectrum 87
 semantic triangle 333
 Semantic Web 91
 semasiological approach 27,
 31, 32
 semi-automatic information
 extraction 301
 set 465
 SGNS 318
 signified 203, 204, 205, 212
 signifier 199, 203, 204, 207, 210
 similarity measure 314
 Simple Knowledge Organisation
 System 94
 single-word term 240
 Singular Value Decomposition
 314, 317
 Sinnform 41, 49, 50
 situated cognition 440
 situated understanding 335
 situatedness 334

- Sketch Engine 298
 sketch grammar 298, 364,
 365, 366
 skip-gram 318
 skip-gram with negative sampling
 318
 SKOS 94
 SNOMED 472
 social variation 405, 409
 socially distributed understanding
 350
 socio-cultural factor 391, 392,
 396
 Sociocognitive Terminology,
 Sociocognitive Theory of
 Terminology 16, 26, 179,
 134, 331, 426
 Sociocognitive Terminology
 studies 331, 336, 338
 sociocultural situatedness 392
 sociolectal variation 409
 Sociolinguistics 178, 183
 Socioterminology 23, 24, 25,
 177, 178, 179, 187, 189, 436
 Soll-Norm 33, 42, 48, 52, 59
 spatial proximity 314
 special terminology 42, 43,
 46, 54
 specialised communication,
 specialized communication
 15, 18, 25, 403, 406, 435, 436,
 445, 454
 specialised diachrony 427,
 429, 434
 specialized concept 403
 specialized corpus 219, 223,
 233, 311, 319, 324
 specialised language discourse
 383, 394
 Specialised Translation,
 Specialized Translation 16,
 19, 24, 29, 30
 specialized discourse 193
 specialized knowledge
 categorization 437, 449
 specialized knowledge
 structuring 437, 439
 specialized language 41
 specialized text 405
 specific concept 71
 spezielle Sprachwissenschaft 46
 spezielle Terminologielehre 43
 sphere of activity 183, 191
 sphere of Terminology 458
 spinning 3D-plots 322
 spontaneous metaphor 386
 standard 87
 standardization 16, 19, 31, 33,
 38, 40, 43, 49, 51, 332
 statistical co-occurrence 313
 status language planning 25
 structural approach to
 terminological growth 466
 structural metaphor 380, 381,
 389, 390
 structure 253, 254
 stylistic variation 406
 subject field knowledge 114
 subject-field expert 114
 subject specialist 461
 subject-specific corpus 91
 subordinate concept 72, 130,
 135
 superordinate concept 72, 74,
 130, 135
 SVD 314, 317
 symbolic approach 231
 synchronic approach to
 terminological growth 422
 synchronic scale 472
 synchrony 41, 44, 421, 423, 434
 synonym 117, 445, 451
 synonymy 15, 33, 34, 48, 59, 50,
 51, 57, 58, 114, 253, 255, 400,
 402, 436, 443, 450
 syntactic co-occurrence 313
 syntactic dependency 315
 syntagmatic relation 253
 syntax-based semantic model
 315
 systematic analysis 117
 systematic terminological
 vocabulary 66
 systematicity 462, 463
 T
 target 213
 target language equivalent 117
 taxonomic relation 241
 taxonomy 87, 253
 taxonomy extraction 321
 TBX 98, 156, 158
 TDC 88
 technical discourse 460
 technical standardization 177
 technolact 25
 TEI 93
 template 356, 361, 374
 temporal relation 69, 71, 77
 tenor 272
 TEnT 91
 term 112, 116, 198, 199, 204,
 205, 206, 210, 240, 241, 249,
 408, 410
 term bank 30, 112, 119, 155, 157
 term extraction 265
 term formation 458, 461
 term formation pattern 468
 term-in-context, terms in context
 117, 272
 term length 470
 term obsolescence 428, 429
 term-oriented corpus analysis
 297
 term record 117
 term variant 168, 169, 193, 404,
 408, 438, 449, 483, 490
 term variation 181, 221, 323,
 435, 439, 443, 445, 446, 484
 term variation research 360,
 436
 Termbase eXchange 98
 Termcat 171
 termhood 92, 470
 TermineSp 171
 termino-ontological resource
 223
 terminographer 27
 terminographic resource 477,
 479, 488, 496
 terminography 22, 30, 32,
 161, 165
 terminological analysis 264, 266
 terminological choice 447, 451
 terminological concept
 management 112
 terminological data 111, 120
 terminological data collection
 88
 terminological definition
 353, 354, 358, 368
 terminological equivalence
 478, 479, 487

- terminological implementation, terminological implantation 25
- terminological growth 458, 461, 468, 472
- terminological knowledge 421
- terminological knowledge base 63, 66, 123, 129, 135, 296
- terminological lexicography 43
- Terminological Markup Framework 98
- terminological metaphor 378, 388, 391
- terminological product 64, 66
- terminological relation 254
- terminological research 264, 275, 279
- terminological resource, terminology resource 66, 87, 88
- terminological sphere 461, 465
- terminological status 240, 246, 249, 253
- terminological unit 112
- terminological variation 274, 283, 312, 319, 323, 399, 401, 402, 415, 435
- Terminologiekennntnis 43
- Terminologiekunde 43
- Terminologielehre 19, 38, 42, 43, 46, 57
- terminologies 459
- Terminologische Grundsatzlehre 52
- terminologist 461
- terminology, Terminology 37, 38, 39, 41, 42, 43, 46, 111, 121, 237
- terminology database 20, 27
- terminology extraction 312, 319, 320, 321, 326
- terminology graph 471
- terminology institute 16
- terminology management system 98
- terminology network 471
- terminology of food 349
- terminology research 436
- terminology science 17, 31, 111
- Terminology theory, theory of Terminology 15, 16, 20, 23, 26, 29, 30, 42, 50, 55
- terminology work 45, 49, 112
- TerminotecaRDF 171
- Termium Plus 157
- termontography 161
- Termonology 26
- text corpus, textual corpus 92, 264
- Text Encoding Initiative 93
- text type 264, 266, 268, 271
- textual genre 223, 263, 264
- Textual Terminology 179, 219, 266, 426
- thematic terminology 92, 113
- thematic terminology work 113
- theoretician 16
- theory-constitutive metaphor 385
- thesaurus 88
- tightly coupled resource 91
- TKB 129
- TM 98
- TMF 98, 156
- TMS 98
- token-level 312
- tool-function relation 117
- top-down analysis 225
- total equivalence 482, 483, 487
- transfer of specialized knowledge 437, 441
- transitional area 389
- Translation Environment Tool 91
- translation equivalence 29, 478, 479, 487
- translation error typology 101
- translation memory 98
- translation quality metric 102
- Translation Studies 15, 29, 265, 283
- transparency 462
- type-level 316
- typology of term variation 442
- U
- UML 93
- unambiguousness 58
- uncertainty 300, 307
- understanding 333, 334
- undirected graph 471
- Unified Modeling Language 93
- unit of understanding 331, 337
- unithood 99
- univocity 52
- unseen item 468
- updating 118
- usage 186
- user interface 123
- V
- vagueness 335
- variation 24, 27, 28, 50, 51, 183, 199, 335
- variation with cognitive consequences 442
- variation without cognitive consequences 442
- variationist theory of terminology 240
- vector-based model 317
- Vector Space Model 316
- verb 240, 246, 249
- verbal economy 451, 452
- verbal interaction 183
- vertex 471
- Vienna Circle 17
- Vienna School of Terminology 39, 41
- visual knowledge pattern 372
- VKP 372, 374
- vocabulary 23, 32, 41, 44, 45, 458
- VSM 316ke
- vulgarization 25
- W
- W3C 93, 151, 170
- W3C DCAT 107
- Web Ontology Language 103
- whole-part relation 117
- Wiener Kreis 39
- window-based semantic model 315
- word-based semantic model 315
- word embeddings 317, 318
- word-oriented resource 91
- Word2Vec 318
- Wording 40
- WordNet 92
- World Wide Web Consortium 93
- Wüster 37, 41, 42, 52, 57, 59
- X
- XML 93

The aim of this volume is to provide an overview of different theoretical perspectives on Terminology, from Wüster to other initiatives that have emerged since the beginning of the 1990s. The volume also covers important topics which have significantly influenced Terminology and its evolution. These include variation, multidimensionality, conceptual relations, and equivalence, among others. The twenty-two chapters of the volume, all written by acknowledged experts in the field, explore the questions that different approaches seek to answer. They also describe the theoretical and methodological principles that were devised over the years to characterize, analyze, and represent terminological data. The semi-chronological, semi-thematic organization of chapters not only provides readers with a clear vision of the evolution of ideas in Terminology, but also gives them an understanding as to why some of these ideas were initially challenged. In addition to being accessible to readers unfamiliar with the basic theoretical principles in the field, the chapters provide a showcase of current research in the field, the challenges looming on the horizon, and finally future directions in terminological research. By bringing together work that is often disseminated in different forums and written in different languages, this volume provides a unique opportunity to look at how different theoretical approaches to Terminology offer complementary perspectives on terms, concepts and specialized knowledge, and help to further a better understanding of the complex phenomena that terminologists must successfully deal with in their work.

ISBN 978 90 272 1106 4



John Benjamins Publishing Company