DE GRUYTER

STUDIES ON MULTILINGUAL LEXICOGRAPHY

Edited by María José Domínguez Vázquez, Mónica Mirazo Balsa, Carlos Valcárcel Riveiro

LEXICOGRAPHICA. SERIES MAIOR

Studies on Multilingual Lexicography

LEXICOGRAPHICA
Series Maior

Supplementary Volumes to the International Annual for Lexicography
Suppléments à la Revue Internationale de Lexicographie
Supplementbände zum Internationalen Jahrbuch für Lexikographie

Edited by
Rufus Hjalmar Gouws, Ulrich Heid, Thomas Herbst,
Anja Lobenstein-Reichmann, Oskar Reichmann,
Stefan J. Schierholz and Wolfgang Schweickard

Volume 157

Studies on Multilingual Lexicography

Edited by
María José Domínguez Vázquez, Mónica Mirazo Balsa
and Carlos Valcárcel Riveiro

DE GRUYTER

ISBN 978-3-11-060467-2
e-ISBN (PDF) 978-3-11-060765-9
e-ISBN (EPUB) 978-3-11-060727-7
ISSN 0175-9264

Library of Congress Control Number: 2019945611

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

María José Domínguez Vázquez, Mónica Mirazo Balsa and Carlos Valcárcel Riveiro

Studies on multilingual lexicography: an introduction

Given the new technological advances and their influence and imprint in the design and development of dictionaries and lexicographic resources, it seems important to put together a series of publications that address this new situation, dealing in particular with multilingual and electronic lexicography in an increasingly digital, multilingual and multicultural society. This is the main objective of this volume, which is structured in two central aspects. In the first of them –**I. Multilingual electronic lexicography in a new society**– the concept of multilingual lexicography is discussed in regard to the influence that the Internet and the application of digital technologies have exercised and continue to exercise both in the conception and design of dictionaries and new lexicographic application tools as well as the emergence of new types of users and forms of consultation. The role of the dictionary must necessarily be related to social development and changes. In the second thematic section –**II. Multilingual electronic dictionaries: projects and tools**–, different dictionaries and resources that focus on a multilingual and electronic approach to the linguistic data for their lexicographical treatment and consultation are presented.

Below are brief summaries that assess the key features of the articles and sections in the volume:

Section I., on electronic lexicography in a new society, opens with *Towards a New Definition of Multilingual Lexicography in the Era of the Internet* by **Pedro A. Fuertes-Olivera** and **Henning Bergenholtz**. The authors address a proposal to differentiate between a traditional and new concept of multilingual lexicography. In order to illustrate the first, Fuertes-Olivera and Bergenholtz describe resources such as *IATE*, *The Logos Dictionary* and the *Diccionari de la Llengua Catalana Multilingüe*. On this basis, they individuate the features of what they refer to as traditional multilingual lexicography. Taking into account that the authors conceive the dictionary not as a product, but as a service, they define the **multilingual dictionaries**, therefore, as a flexible and integrated information tool, and propose a definition of multilingual lexicography, “as the theory and practice of unified and well-connected monolingual, bilingual and multilingual dictionaries using data from a multilingual database”. The *accounting dictionaries* serve to exemplify their proposals.

The study by **Rufus H. Gouws** –*Metalexicographic models for multilingual online dictionaries in emerging e-societies*– addresses the challenges posed for lexicography by the digital age, not only in the planning and compilation of dictionaries but in relation to new types of users with their respective needs and abilities. Gouws’

https://doi.org/10.1515/9783110607659-001

work begins with an exemplification of the development of the lexicographic practice, using examples from the English-speaking lexicographic environment, which testify to the different role that the dictionary has had in society. As an example, for an emerging multilingual and multicultural e-society with a new generation of potential dictionary users and user's needs, Gouws describes the South African lexicographic environment. After discussing the South African case, a pilot project, MobiLex (a web-mobile application that allows users to search for terms from certain subject fields), is presented. In this way, this contribution shows how dictionaries can solve not only linguistic but also cultural loopholes. The author also focuses on the need to promote the establishment of a dictionary culture. For his part, **Sven Tarp**, in his contribution *A dangerous cocktail: databases, information techniques and lack of visions*, addresses the challenges for lexicography in the development of information technology and new technologies. The author provides examples of how to deal with them through the design and implementation of databases, user interfaces or other tools that would help in compiling and presenting information from online dictionaries. In this way, Tarp argues for the need to open new research centred on the importance of empirical studies for the production of dictionaries and advocates the development of a lexicographic theory and a methodology in which lexicography and new technologies are integrated in such a way that effective solutions can be offered to all problems.

Section II of this volume focuses on the presentation of multilingual projects and tools. In *Multilingual Electronic Dictionary of Motion Verbs (DICEMTO)*: overall structure and the case of andar, **Olga Batiukova** and **Elena de Miguel** present DICEMTO, an electronic-format multilingual dictionary of verbs of movement. Its main objective is the systematic registration of the different senses adopted by these verbs depending on context and as a result of their combination with other features of the predicate. As a sample, Batiukova and de Miguel analyse in detail the lexical entry of the verb *andar*, composed by the minimal definition, argument structure and thematic structure, event structure and qualia structure, before proceeding to point out the semantic variations produced by the interaction of different meaning components of the different levels. The contribution *From the Linguaturismo glossary to the Dictionary of Food and Nutrition: proposal for a new electronic multilingual lexicography* by **Maria Vittoria Calvi** and **Luis Javier Santos López**, details a proposal for a new electronic multilingual lexicography centred on an approach based on textual genres that, according to the authors, would be very useful both in the field of terminology and lexicography. As an example of the application of this methodology, the project Linguaturismo, a Spanish-Italian bilingual terminology glossary, has been developed on the basis of comparable textual corpora of both languages and whose main objective is the systematic analysis of genres used in the field of tourism. The article then addresses the *Dictionary of Food and Nutrition*, a specialized dictionary with a multilingual and polyalphabetic terminology covering Arabic, Chinese, English, French, German, Italian, Portuguese, Spanish and Russian. The possibilities offered by corpus linguistics and the genre-based approach would therefore be, according to the authors, very useful for an innovative and effective development of terminological and lexicographical products. The article by **Gloria Corpas Pastor** and **Isabel Durán-Muñoz** *–INTELITERM: In search of efficient terminology lookup tools for translators–* describes an innovative tool that considers the translator as its main recipient. This resource combines corpus management tools with different search and customization options in order to optimize translation results and minimize the effort of the translator in the terminological search. After setting out the topic, the authors undertake a detailed analysis of this tool, which together with the functions of assistance in translation, also offers a database editor. Furthermore, they provide empirical data on the degree of user satisfaction with INTELITERM. A multilingual and polydirectional dictionary is the focus of the study by **María José Domínguez Vázquez** and **Carlos Valcárcel Riveiro**, *PORTLEX as a multilingual and cross-lingual online dictionary*. This dictionary, an accessible reference tool for a wide range of users with different lexicographical needs, includes learning and teaching languages, translation, grammatical research and natural language processing (NLP). The contribution by Domínguez and Valcárcel assesses the main characteristics of this electronic multilingual and valential portal. The authors also explain the different stages of the portal's development, the structure of its database and the user interface, detailing different types of search. Different possibilities of interaction and collaboration as well as the creation of a virtual community of users and editors around PORTLEX are also analysed in this study. The needs of translators in relation to terminology management and consultation during the different phases of translation tasks is the focus of the contribution by **Isabel Durán-Muñoz** and **Gloria Corpas Pastor**, *Corpus-based multilingual lexicographic resources for translators: An overview*. This study provides a broad overview of different types of multilingual electronic resources that allow different types of comparable or parallel corpora access, which typically use the Web as Corpus (WaC). Noting the usefulness of these resources, the authors also address some of their shortcomings and come to the conclusion that it is imperative to implement tools specifically aimed at translators. In this line of integration of corpora in electronic dictionaries, current research projects are presented that would cover the needs of translators and that would allow formulating, implementing and evaluating new options for management and terminological acquisition. In *Construction of a WordNet-based multilingual lexical ontology for Galician*, **Xavier Gómez Guinovart** and **Miguel Anxo Solla Portela** present some aspects of the development of Galnet, a WordNet-based multilingual lexical ontology for Galician that aims to build a WordNet for Galician aligned with the ILI generated from the English WordNet 3.0, following the expand model (Vossen 2002) for the creation of new wordnets. Following a general description of WordNet, the paper focusses on wordnet's applications in the fields of terminology acquisition and ontology learning and management. In the first case, the application in terminology acquisition, the authors explain new strategies for exploring terminology in the lexical-semantic network as they describe the Termonet tool. In the second case, the applications in ontology learning and management, their paper examines examples and procedures for applications in WordNet semantic areas –particularly the relationship between synsets and epinonyms– and the applications in the Semantic Web –for example, providing the contents of Galnet as a Resource Description Framework. In *Designing and compiling a terminological and multilingual dictionary for language teaching and learning: key issues and some reflections*, **Ignacio M. Palacios** and **Mario Cal** provide an overview of the process of developing a multilingual terminology dictionary for language teaching and learning. The authors justify, first of all, the need for a reference work such as this one and describe the typology of users who are intended as its recipients, before focusing on the content and architecture of the information offered. Palacios and Cal also present different possibilities of future developments such as the development of a user-friendly online version, which would have user feedback and that would be complemented with the implementation of other languages. The rapid growth experienced by the football world owing to its social impact, mediatisation and marketing is the starting point for the lexicographic conception of a multilingual (Polish, Russian, English and German) hard copy LSP dictionary of football terms presented by **Janusz Taborek** in *Multilingual LSP dictionary. Lexicographic conception of a dictionary of football language*. The dictionary, with more than 2400 entries in each language, follows the author's previous lexicographical project, the German-Polish language dictionary (Taborek 2006) and aims to provide a solution to a number of problematic issues that arise at different levels. In this way, Taborek describes the macrostructure of the dictionary and discusses the problems that may be posed by the lemmatization or the multi-word units, the microstructure, addressing issues such as markers or examples based on use, and, finally, the structure and reference system employed.

The monograph, in short, includes different works following a very specific thread, such as the multilingual lexicographic approach to phenomena and linguistic realities. It is also a clear testimony to the need to carry out studies of these characteristics, since an inexhaustible source of information can be extracted from them, which will undoubtedly have an impact on the optimization of existing resources and on the design of new tools. We believe, therefore, that the target readership may be researchers and teachers, but also future researchers in search of a thematic line of study.

We cannot end this introduction without expressing our gratitude to the Spanish Ministry of Economy and Finance, which supported the PORTLEX project (FFI2012-32456) from where the initiative to produce this volume began, to the BBVA Foundation's funding for the *MultiGenera* project (Foundation Grants for Scientific Research Teams 2017) and to the Spanish Ministry of Economy, Industry and Competitiveness/Spanish State Research Agency, as well as to the European Regional Development Fund, for co-financing the MultiComb project (Scientific and technical excellence research program, FFI2017-82454-P)[1]. We should also acknowledge the continued support from the Galician Lexicography Network RELEX (ED341D R2016 / 046). Finally, we would like to thank all those who have participated unselfishly in the process of evaluation –review and revision–, as well as the authors who with their contributions have made this work possible.

[1] Both projects, MultiGenera and MultiComb, develop new tools on the basis of PORTLEX.

Section 1: Multilingual electronic lexicography in a new society

Pedro A. Fuertes-Olivera and Henning Bergenholtz

Towards a new definition of multilingual lexicography in the era of internet

Abstract: This chapter reviews the traditional concept of multilingual lexicography and offers some examples of traditionally conceived multilingual lexicographic projects that are abundant in the era of the Internet: IATE; The Logos Dictionary; The Diccionari de la Llengua Catalana Multilingüe. Our analysis shows that this traditional approach is erroneous and does not help potential users. Consequently, we offer a different approach, which stems from our definition of multilingual lexicography as the theory and practice of unified and well-connected monolingual, bilingual and multilingual dictionaries using data from a multilingual database. These dictionaries are information tools that cover words, terms, facts, and/or things in several languages, have the same conceptualization at the pre-compilation phase, and make use of lexicographic and technological know-how. This allows (a) lexicographers to add new languages to the same information database from which new monolingual, bilingual and multilingual dictionaries can be extracted, and (b) users to retrieve connected data easily and to spot and understand possible similarities and differences among the several languages covered. We also illustrate the operational side of this definition in some of our accounting dictionaries.

Keywords: multilingual lexicography, e-lexicography, function theory, dictionary writing system, integrated information tools

1 Introduction

Although the coming of age of the Internet has brought about many changes and modifications to lexicography, it has not affected the *core of lexicography*, i.e. the aspects and elements common to all dictionaries and other lexicographic reference works regardless of their nomenclature (Fuertes-Olivera/Bergenholtz 2011, Fuertes-

Pedro A. Fuertes-Olivera: International Centre for Lexicography, University of Valladolid (Spain), Department of Afrikaans and Dutch, University of Stellenbosh (South Africa), Plaza del Campus 1, 47011, Valladolid, tel. +34 983423582, pedro@emp.uva.es
Henning Bergenholtz: Centre for Lexicography, University of Aarhus (Denmark), Department of Afrikaans and Dutch, University of Stellenbosh (South Africa), Department of Information Science, University of Pretoria (South Africa), College of Foreign Studies, University of Jinan (China), Centre for Lexicography, School of Communication and Culture, Fuglesangs Allé 4, 8210 Aarhus V (Denmark), hb@bcom.au.dk

https://doi.org/10.1515/9783110607659-002

Olivera/Tarp 2014, Tarp 2008[1]). These are information tools that are designed and compiled for assisting specific types of users to satisfy their punctual information needs in specific types of extra-lexicographic situations in a quick and easy way. This has several implications, but one that is relevant for this chapter is the idea that users, lexicographic data and access routes –the three core lexicographic elements– are interconnected, and, therefore, whatever modification is made in any of them must be approached holistically, both in theory and in practice.

A holistic approach basically means that any important change in, let us say, the way lexicographic data are searched for may also affect target users and access routes. For instance, a lexicographic project using literary works as source data and targeting native speakers does not need links to external corpus data. The merits of links to such data may not compensate for the amount of time and resources needed for preparing the corpus, selecting the concordances, paragraphs, hyperlinks and so on, especially as most target users –native speakers in this hypothetical case– require only a few examples and contexts of use for disambiguating the meaning and usage of most lemmas.

Lexicographic modifications and changes are more typically connected with the *specifics of lexicography*, i.e. the lexicographic features that make a particular dictionary or reference work different from others. The specifics of lexicography highlight the interdisciplinary vocation of lexicography and its connections with many other different disciplines. This interdisciplinary vocation also explains the necessity of adopting a holistic approach to dictionary making. For instance, the conception of an online dictionary cannot take place without IT knowledge (Bergenholtz/Nielsen 2013). Similarly, crafting definitions of, for instance, accounting terms is inconceivable without accounting knowledge (Fuertes-Olivera/Tarp 2014). Many dictionary projects fail or result in poor quality reference works (e.g. *IATE*) because they do not adopt a holistic approach, i.e. one in which user's needs as well as lexicographic data and access routes to these data are *conceptualised* holistically and evaluated in terms of cost and time.

This chapter proposes a modification of the concept of multilingual lexicography as this is traditionally understood and used (Section 2). Our proposal respects the core and specifics of lexicography in the era of the Internet, employs some of the technical possibilities *idiosyncratic lexicographic databases* (i.e. storage systems specially prepared and designed for a particular lexicographic project) offer, and provides an illustration of our work in several lexicographic projects; in these we have prepared an array of changes and modifications, all of which take into consideration the holistic relationship among the three main components of lexicography (Section 3). A final conclusion summarises the main ideas discussed (Section 4).

[1] Thanks are due to the Ministerio de Economía y Competitividad for financial support (Grant Ref. FFI2014-52462-P).

2 Multilingual lexicography: traditional view

In traditional lexicographic thinking, monolingual lexicography is understood as the part (of lexicography) dealing with the theory and practice of dictionaries covering one language. Similarly, bilingual lexicography is concerned with the theory and practice of dictionaries which incorporate two languages, and multilingual lexicography with these criteria in dictionaries covering three or more languages. In this traditional view, a dictionary is apparently only about words. We believe that this view is a fallacy and, therefore, we advocate and adopt a broader view of *dictionary,* i.e. a lexicographic tool that deals with words, terms, facts and things. However, for the purpose of this chapter we will use language or words as an umbrella term for these four items.

Traditional lexicographers also describe several subtypes of monolingual, bilingual or multilingual dictionaries, which are grouped in terms of the number of languages used. Regarding monolingual lexicography, lexicographers typically use the lemma language L_1 (= A) and the explanation language in L_1 (in most cases), or in L_2 for dictionary users with L_2 as the mother language (Figure 1):

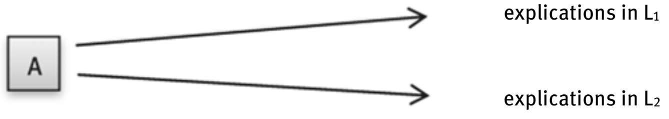


Fig. 1: Typical Monolingual Dictionary

There may be some differences, such as in particular the presence of variants of A in the lemma list of certain dictionaries, e.g. a dictionary of Spanish in which the lemma list also includes variants of Mexican Spanish, Cuban Spanish, or of some dialect, e.g. Andalusian. Figure 2 shows this sub-type of monolingual dictionaries:

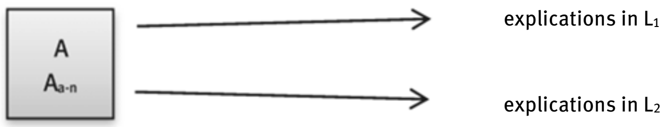


Fig. 2: Monolingual Dictionaries with Lemma Variants

Regarding bilingual lexicography, the situation is similar, although here there are more models or subtypes of dictionaries. Firstly, the lemma list of unidirectional dictionaries contains lemmas from L_1 (= A) and the equivalent list contains explanations in L_1, L_2 (= B) or in both L_1 and L_2 (Figure 3):

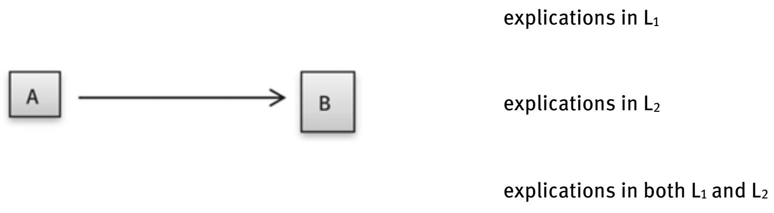


Fig. 3: Unidirectional Bilingual Dictionaries

Secondly, if the lemma list contains lemmas with variants of L1, lexicographers also include variants of A (Figure 4):

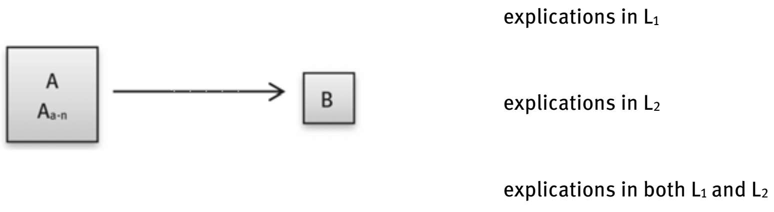


Fig. 4: Unidirectional Bilingual Dictionaries with Variants of Lemmas

Thirdly, in very few cases, the equivalent list also contains variants of B (Figure 5):

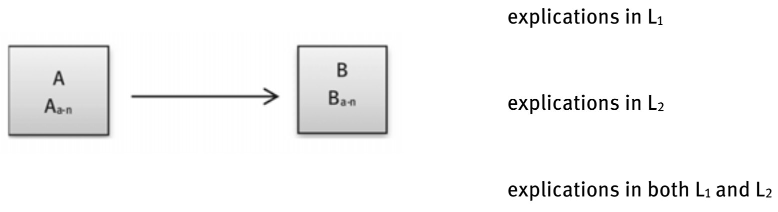


Fig. 5: Unidirectional Bilingual Dictionaries with Variants of Lemmas and Equivalents

Fourthly, the lemma list of bidirectional dictionaries contains lemmas from L_1 (= A) and the equivalent list from L_2 (= B) with explications in both directions in L_1, L_2 or in both L_1 and L_2 (Figure 6):

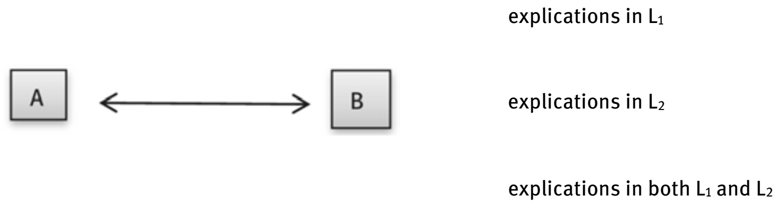


Fig. 6: Bidirectional Bilingual Dictionaries

Finally, variants of both A and B can also be included in the lemma list (Figure 7):

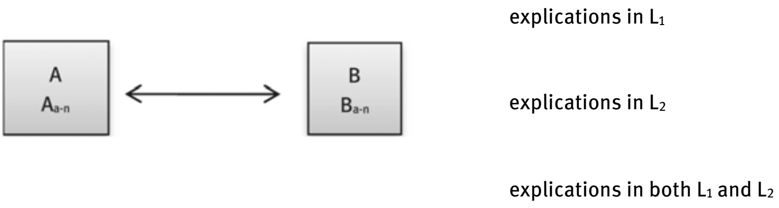


Fig. 7: Bidirectional Bilingual Dictionaries with Variants of Lemmas and Equivalents

Regarding multilingual lexicography, the situation is usually simpler: there is a lemma list of A (with or without variants of A) and two or more equivalent lists (B, C, D, E, etc.) also with or without variants (Figures 8 and 9). Items of meaning, collocations, examples, grammar data types and so on are not normally included:

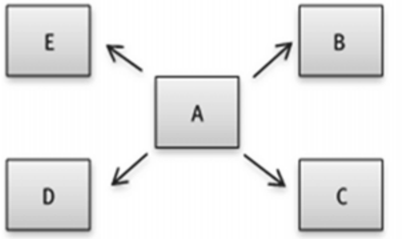


Fig. 8: Typical Multilingual Dictionaries

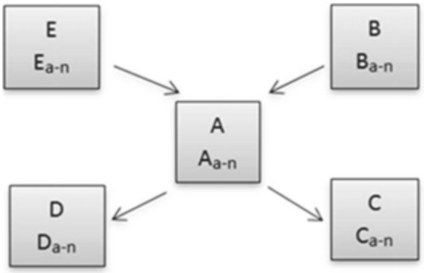


Fig. 9: Multilingual Dictionaries with variants of Lemmas and Equivalents

The above models (represented as Figure 1 to 9) are widespread and can be easily illustrated with examples taken from either print or online lexicographic projects. For the purpose of this chapter, we will focus on three online lexicographic projects that represent widespread practice in traditionally-conceived lexicographic circles:

(a) *InterActive Terminology for Europe* (*IATE*), (b) *The Logos Dictionary* and (c) the *Diccionari de la Llengua Catalana Multilingüe*.

2.1 InterActive Terminology for Europe (IATE)

On the homepage of the *Inter-Active Terminology for Europe* (*IATE*), we are informed that *IATE* is the "EU inter-institutional terminology database". An analysis of publications and Internet documents on *IATE* shows that it has been used in EU institutions and agencies since 2004 for the collection, dissemination and management of EU-specific terminology. It is the successor of several EU databases: EURODICAUTOM[2], TIS, EUTERPE, EUROTERMS, and CDCTERM. *IATE* was launched "with the objective of providing a web-based infrastructure for all EU terminology resources, enhancing the availability and standardisation of the information"[3]. In March 2007, it provides the general public with free access to multilingual terminology in the fields of activity of the European Union.

For the purpose of this chapter, *IATE* is an example of a multilingual or plurilingual translation-oriented specialised online dictionary, the main characteristics of which are summarised below (our analysis):

- *IATE* covers a broad spectrum of domains: politics, international relations, European Communities, law, economics, trade, finance, social questions, education and communications, science, business and competition, employment and working conditions, transport, environment, agriculture, forestry and fisheries, agri-foodstuffs, production and technology research, energy, industry, geography and international organisations. Each record typically stores the information supplied by EU sources, which has resulted in an information system that contains more than 1.6 million records, around 7 million terminological entries, and 8.7 million terms, "including approximately 540,000 abbreviations and 130 phrases"[4].
- *IATE* covers 25 official languages; English, French, and German being the languages with more terms. For instance, there are about 1.6 million English terms and only around 35,000 in Hungarian. In addition, it includes Latin for botanical and zoological names.

[2] See at [<http://iate.europa.eu>; last access: April 17, 2017].

[3] See at [<http://iate.europa.eu>; last access: April 17, 2017].

[4] See at [<http://iate.europa.eu>; last access: April 17, 2017].

- *IATE*’s genuine purpose(s) is to facilitate translation and terminology processing, which is considered essential for supporting translation, e.g. in multilingual organizations *(UNTERM*) and multilingual countries (Canada). Hence, *IATE* mainly targets translators, interpreters and writers of EU texts who access data aimed at increasing the reliability of the solution proposed in all (or many) of the EU languages.
- *IATE* records have three levels. At the top level, users access language-independent data, whilst the second contains language data and the third includes term data. This means that *IATE* is based on concepts and not merely terms, offering more definitions, relevant references, fewer duplicates, and multi-languages entries.
- *IATE* highlights the search string (e.g. *casa* in Example 1) and then offers an array of equivalents in all or some EU languages. For instance, there are 13 equivalents for Spanish *casa* in the domain “Urbanismo y Construcción” (Construction and Town Planning) (Example 1):

Ex. 1: Entry of casa in IATE

ES	**casa**
DA	hus
DE	Haus
EL	οıkία
EN	house
FI	talo
FR	maison
LA	maison
domus	
NL	vrijstaand huis
vrijstaande woning	
PT	casa
habitação	
SV	hus

Clicking on any of these equivalents will recover more information, e.g. a brief definition of Danish *hus*, its source, domain, date of inclusion and reliability (Example 2):

Ex. 2: Entry of casa in IATE

Domain	Construction and town planning
Definition	man skelner mellen enfamiliehuse
og etageboliger	
Definition Ref.	Bendtsen, Byplanlægning 1969
Term	**hus**
Reliability	3 (Reliable)
Term Ref.	Bendtsen, Byplanlægning 1969
Date	24/09/2003

We believe that *IATE* could be improved provided several of the following drawbacks are addressed and resolved:

a) It assumes that multilingual dictionaries consist of monolingual dictionaries that are "joined", i.e. each dictionary has its own lexicographic database, is compiled separately, and then juxtaposed with the aim of offering data in a fixed and static order: data in language A, data in language B, data in language C and so on. In other words, a multilingual dictionary such as *IATE* is a collection of monolingual dictionaries that are displayed in a fixed order for all users and all-use situations.

b) It assumes that multilingual dictionaries are basically concerned with including the translation equivalents of the search string.

c) It assumes that user's needs and quick and easy access are unnecessary features and that therefore users need several clicks for disambiguating meanings. In addition, contexts of usage are not usually included.

2.2 The Logos Dictionary

The *Logos Dictionary* is a lexicographic product that can be accessed at Logos homepage[5]. Logos is a company that "delivers translation solutions with more than 35 years of knowledge and experience" (Homepage). The *Logos Dictionary* represents a well-known lexicographic product in the era of the Internet, one of its most

[5] See at [<http://www.logosdictionary.org>; last access: April 17, 2017].

defining characteristics being that compilers pay much more attention to technological gadgets than to lexicographic theory, as shown below (Example 3):

Ex. 3: The entry house in The Logos Dictionary (Excerpts)

house Noun

[Subject: Architecture]

A structure serving as a dwelling for one or more persons

Translated by: ROBERT GRIGOR BAXTER

Synonym: home

Translated in 168 languages

Translations

Italian	**casa**; abitazione
Spanish	**casa**; techo; residencia; vivienda
French	**maison**; chez-soi; domicile; logis
	→(more translations up to 168 languages)

Our analysis indicates that this dictionary has similar drawbacks to those of the IATE. In addition, it is also interesting to highlight three features of lexicographic projects such as the *Logos Dictionary*:

- It is a collaborative project. This has three main lexicographic implications. The first is that no lexicographer has in fact been (or is) in charge or the project. In other words, its conceptualisation at the pre-compilation phase was not undertaken by lexicographers. The second is that its data types are constantly growing, most typically by additions and corrections made by (freelance) translators working for *Logos*. For instance, at the time of writing this chapter (February, 2016), it has almost 7,600,000 entries. This impressive figure cannot be evaluated as the dictionary covers more than 100 different languages with very different ways of dealing with dictionary entries. To sum up, its dictionary data stock and lexicographic treatment is subject to many "on-the-spot" decisions, with no one in charge of the project. Finally, it is a "freely-accessible" dictionary, which also means that no one is really responsible for its content and functioning system.
- It mainly targets translators and writers of professional texts. However, it offers data that translators do not need (e.g. translators' names), but does not contain much necessary data, e.g. translated contexts of use and other relevant grammar data.
- The access system is very complicated and unsystematic. For instance, the tag "subject" contains almost 200 possibilities, (some of which are "Africa", "Arts", "Algebra", etc.).

2.3 The Diccionari de la Llengua Catalana Multilingüe

The *Diccionari de la Llengua Catalana Multilingüe* is a project compiled at the Universitat Pompeu Fabra (Barcelona) which contains two different parts. The first is a monolingual Catalan dictionary, accessed by writing a word in the search engine and clicking on the bottom "catalá" at the top of the homepage. It contains around 20,000 dictionary entries and up to 30,000 meanings. The second part consists of Spanish, English, French and German equivalents of the Catalan word, some of which are accompanied with a Catalan semantic tag in brackets (Example 4 of Catalan *casa*):

Ex. 4: The entry house in The Logos Dictionary (Excerpts)

casa
F1 Nom genèric de tot edifici destinat a servir d'habitació humana.
2 Edifici o part d'un edifici en què hom habita amb els seus. Anar a casa dels pares.
3 a casa (de) LOC PREP Són a casa l'àvia. Són a casa de la Joana.
4 anar vestit d'estar per casa Anar vestit d'una manera senzilla, descurada.

Castellà: casa; (edifici, institució) casa; (família, llinatge) casa; **a casa (de)** en casa de; **casa gran(o de la vila)** ayuntamiento, casa de la villa; **anar vestit d'estar per casa** llevar un vestido de andar por casa, estar de casa
Anglès: house; (llar) home; (família) household; (edifici) Building**; casa de la vila** town hall; **casa e pisos** block of flats; **casa de pagès** country House**; a casa** (at) home**; tirar la casa per la finestra** to spare no expense; **vaig vestit d'anar per casa** I'm in my ordinary clothes
Francès: (edifici) maison; (habitatge) maison, domicili M, demeure; (família) famille; (llinatge) maison**; a casa de** ... chez... A casa meva, chez moi**; d'estar per casa** (descurat) négligé –e
Alemany: Haus N; Una mentida com una casa, ee Rieseniüge; **a casa** zu Hause; **a casa de** bei; **casa de la vila** Rathaus; **d'estar per casa** anspruchslos, bescheiden

In addition to what we have commented on for the *IATE* and the *Logos Dictionary*, the main feature of the *Diccionari de la Llengua Catalana Multilingüe* is that its data types are not connected, which makes users doubt, for instance, the relationship between meaning entries and equivalents and between the different equivalents. For example, Spanish *casa* as an institution (*institució*) does not have an equivalent in the rest of the languages included. Similarly, the phrases are different and have different lexicographic treatment. For instance, *town hall* (*ayuntamiento* in Castellá*)* is in Catalan *casa gran* or *casa de la vila*; in English, the Catalan equivalent of *town hall* is *casa de la vila* but not *casa gran*.

To sum up, our previous analysis shows that the traditional concept of multilingual lexicography rests on very shaky and unstable foundations, such as:

a) Devising multilingual dictionary projects lacks suitable conceptualisation at the pre-compilation phase. These projects mainly consist of juxtaposing monolingual dictionaries, sometimes ones which are very different. For instance, German *Haus* goes with N (grammar data), whereas Spanish *casa*, English *house* and French *maison* do not have any grammar tag (Example 4). In addition, the Spanish, English and French equivalents are accompanied by a semantic tag, but the German one does not have it. In other words, the multilingual dictionary project is devoid of its own proper conceptualization at the pre-compilation phase, which results in an array of unconnected lexicographic data that are simply juxtaposed, which forces users to "imagine" possible connections and relations among them.

b) Multilingual dictionaries typically contain a small number of data types: most, if not all of them, only include lemmas with or without meaning items and translation equivalents without meaning items;

c) They only show static articles.

d) They are not very helpful for most users: they generally include unconnected data types, which obliges users to connect them by themselves.

e) Core lexicographic elements such as data types, users' needs and access routes are not treated in a holistic way. For instance, in most existing online dictionaries, e.g. in the *IATE* and the *Logos* Dictionary, users have to search two or three times for retrieving something other than equivalents without contexts.

3 Towards a new definition of multilingual lexicography in the era of the Internet

The influence of the Internet on lexicography can be felt in several areas. Some of these are connected with the business model of publishing houses, whereas others affect one or more elements associated with the core or the specifics of lexicography. An analysis of these modifications will help us to present a new definition of multilingual lexicography in the era of the Internet.

The business model has changed so much that well-established publishing houses, e.g. Macmillan, have ended the publication of printed dictionaries. Moreover, well- acclaimed projects such as the *Encyclopaedia Britannica* or the *New Palgrave Dictionary of Economics* are undergoing large-scale modification as a result of the developments of projects such as *Wikipedia* and the opportunities the Internet offers. Accordingly, multilingual lexicography in the era of the Internet must completely change its traditional approach for three main business-connected reasons. The first is that freely-accessible multilingual dictionaries such as *IATE, The Logos Dictionary,* and *the Catalan dictionary* are abundant and easy to retrieve from the Internet. Hence, it would be foolish to prepare, let us say, a business-oriented lexicographic project —i.e. a dictionary to be sold— similar to those that can be browsed for free. The second is that the Internet offers more opportunities for niche markets than for mass markets. This means that multilingual lexicography must explore the possibilities of entering the former with a niche product, i.e. a good or service with features that appeal to a particular market subgroup. The third reason is that only companies with Internet-oriented business models can really deal with this new situation, in which around-the-clock call centres and assistance personnel as well as cloud computing facilities are totally necessary.

To take into consideration possible business models in the era of the Internet also demands a clear understanding of the main features of the information tool we are going to sell. For instance, we believe that online dictionaries can no longer be sold as goods but as services. This means that potential consumers do not pay for a finished product but for an on-going service. This has six main lexicographic implications. The first one is that there must always be an appropriate system of communication between users and lexicographers for answering queries, updating entries, introducing new entries, and so on. The second is that systems with easy and simple access routes are a *sine qua non,* as most users reject very sophisticated “advanced search systems”. The third is that lexicographers must not use symbols, abbreviations, convoluted language and so on, as these may complicate the understanding and usage of the lexicographic data searched. The fourth is that we need more data types for describing the meaning and usage of each dictionary entry. Dictionary entries such as those of examples 1, 2, 3, and 4 are of very limited use. The fifth implication is that users need a simple and well-conceived user guide. Finally, dictionary entries must always be well-presented, e.g. by using suitable colours and fonts, meta-language presenting the data type the user has at their disposal, and so on. For instance, a dictionary entry such as example (4) is poorly conceived and its presentation does not help at all: users can only confirm their intuition(s) regarding, let us say, equivalents of Catalan *casa*. In other words, they have to decide by themselves on the possible pairing of the different meaning items of *casa*. Is English *house* or French *maison* the same object? Are they what meaning item 1 of *casa* explains? Or do they correspond to meaning item 2?

We have connected the above reflections with the core and specifics of lexicography with the aim of presenting and illustrating a new definition of multilingual lexicography in the era of the Internet. Our first idea is that *multilingual lexicography deals with dictionaries that have the same conception at the pre-compilation phase*. This means that a multilingual dictionary is not a collection of different dictionaries juxtaposed, but the same dictionary with the same elements for all the languages covered. For instance, the *accounting dictionaries* (Nielsen et al. 2003 and 2006, Fuertes-Olivera et al. 2012 and 2013) have the same conceptualization for the three languages they cover: English, Danish and Spanish. The dictionary entry for *accounts* contains the same key lexicographic data for its Danish and Spanish counterparts: English grammar; an English definition followed by the Danish or Spanish equivalent, their grammar data and contexts of usage (English collocations and an example translated into Danish or Spanish) (Example 5):

Ex. 5: The entry house in The Logos Dictionary (Excerpts)

accounts (English-Danish)
<noun> plural
Definition
The accounts are the detailed records of an enterprise's financial affairs showing the income and expenses for a period as well as the financial position at a particular date.
regnskab
substantiv <et regnskab; regnskabet, regnskaber, regnskaberne>
Collocations

- annual accounts (UK)
- årsregnskaber
- audit the accounts (UK)
- revidere regnskabet
- audited accounts (UK)
- revideret regnskab

(...)

Examples

- These accounts were approved by the Board of Directors on 8 March 2006. (UK)
- Nærværende årsregnskaber blev godkendt af bestyrelsen den 8. marts 2006.

UK

Synonymer
financial statements (IAS/IFRS+US)
accounts (English-Spanish)
<noun> plural
Definition
The accounts are the detailed records of an enterprise's financial affairs showing the income and expenses for a period as well as the financial position at a particular date.
cuentas anuales
<no singular, unas cuentas anuales, las cuentas anuales>
Collocations

- annual accounts (UK)
- cuentas anuales
- audit the accounts (UK)
- auditar las cuentas anuales
- audited accounts (UK)
- cuentas anuales auditadas

(...)

Examples

- These accounts were approved by the Board of Directors on 8 March 2006. (UK)
- El Consejo de Administración aprobó estas cuentas anuales el 8 de marzo de 2006. ()

Synonymer
financial statements (IAS/IFRS+US)

Our second idea is that *multilingual dictionaries must be equipped with singularising lexicographic systems and technologies,* i.e. lexicographic concepts and/or technological options that allow users to find out possible differences among the languages covered. There are some possible ways of putting this idea into practice. Firstly, the editing system can be equipped with technologies that allow lexicographers to include or exclude data. For instance, in the *accounting dictionaries,* a device such as "include" or "exclude" allows lexicographers to determine the amount of data to be retrieved. For instance, if a user searches for *associate* (English) they will recover one English-Danish entry and two English-Spanish entries (Example 6). (This means that English *associate* has one Danish equivalent but two different Spanish ones. Hence lexicographers will associate the same English lemma to different Spanish and Danish data):

Ex. 6: The entry house in The Logos Dictionary (Excerpts)

associate (English-Danish)
Definition
An associate is an enterprise which is not a subsidiary, but in which another enterprise (the parent) and its subsidiaries has a significant influence over the operating and financial policies, typically by holding between 20 and 50 per cent of the voting equity instruments.
associeret virksomhed
associate (English-Spanish)
Definition
An associate is an enterprise which is not a subsidiary, but in which another enterprise (the parent) and its subsidiaries has a significant influence over the operating and financial policies, typically by holding between 20 and 50 per cent of the voting equity instruments.
asociada
An associate is a member of an organisation or professional body.
asociado

Secondly, lexicographers can use tags, lexicographic usage notes or different data for explaining differences. For instance, in the *accounting dictionaries,* tags such as "DK", "UK", "US", "IAS/IFRS" and "E" are used for indicating that the dictionary entry is employed in Danish accounting, UK accounting, USA accounting, International accounting or Spanish accounting, respectively. In addition, there are usage notes for explaining differences among the languages covered. For instance, a note such as "In Spanish legislation, shares can have different benefits but not different political rights" is included in some entries in order to indicate differences between Spanish accounting and the other two accounting languages covered in the *accounting dictionaries.*

Our third idea is that *multilingual dictionaries must be part of a unified lexicographic system with different search and access possibilities.* For example, the *accounting dictionaries* use a hub-and-wheel structure (Nielsen/Almind 2011) consolidated at the level of definitions. In practical terms, and for the purposes of this chapter, this has three main implications. The first is that each of the languages covered can be retrieved individually. For instance, in the *accounting dictionaries* users can only search for English, Danish or Spanish data and recover similar but different data. The English entry *assets* (Example 7) and the Spanish entry *activos* (Example 8) refer to the same entity but the linguistic and factual description is different. Here, in the Spanish entry, users have more collocations, one example and a link to an outer text, although no synonym is included; meanwhile, in the English entry there is a synonym, but no example or link to an outer text:

Ex. 7: The entry *assets* in the English accounting dictionary

assets
Inflexion
noun <plural>
Definition
The totality of what a person or enterprise owns is called his or its assets.
Collocations

– all assets
– total assets
– transfer the assets
– use of the company's assets

Synonym
property

Ex. 8: The entry *activos* in the Spanish accounting dictionary

activos
nombre masculino <no singular; unos activos, los activos>
Definición
Los activos es el conjunto de bienes, propiedades y derechos propiedad de una persona física o jurídica.
Colocaciones

– el conjunto de los activos de una empresa
– la gestión de activos empresariales
– los activos de renta fija
– activos disponibles
– administrar activos
– todos los activos
– transferir los activos
– uso de los activos de la empresa

Ejemplos
- Los activos son un recurso o bien económico propiedad de un negocio, con el cual se obtienen beneficios. Los activos de un negocio varían de acuerdo con la naturaleza de la empresa.

Fuente
Wikipedia
[<http://es.wikipedia.org/wiki/Activo_(contabilidad[6])>]

The second implication is that lemmas and equivalents are *always* united by definition, and therefore connected, which eliminates the possibility of making wrong assumptions and facilitates consultation, even when related entries have different data (Examples 7 and 8, above). In such a system, users do not have to make guesses, as every data type is accompanied by its equivalent data type in the second, third or n-language. For instance, searching, let us say, *assets* (Example 9) will always retrieve Danish and Spanish connected entries (equivalents *formue* and *actives,* respectively):

Ex. 9: The entry *assets* and its Danish and Spanish counterparts

assets (English-Danish):
Definition
The totality of what a person or enterprise owns is called his or its assets.
formue
Definition
Formue er den samlede mængde af, hvad en person eller en virksomhed ejer.
Synonymer
aktiver
kapital
værdier
Collocations
- all assets
- hele formuen
- total assets
- samlet formue

Synonym
Property
assets (English-Spanish)
Definition
The totality of what a person or enterprise owns is called his or its assets.

[6] Last access: April 17, 2017.

activos
Definition
Los activos es el conjunto de bienes, propiedades y derechos propiedad de una persona física o jurídica.
Collocations
- all assets
- todos los activos
- transfer the assets
- transferir los activos

Synonym
Property
Source
Wikipedia

However, searching for Danish *formue* or Spanish *activos* retrieves the same concept in a different context. This occurs because lexicographers select data from different sources. For instance, data for describing *activos* come from Spanish accounting legislation, which is clearly different in the Danish or English case (Example 10):

Ex. 10: The entries *formue* and *activos* in the accounting dictionaries

formue (Danish-English)
Definition
Formue er den samlede mængde af, hvad en person eller en virksomhed ejer.
assets
Definition
The totality of what a person or enterprise owns is called his or its assets.
Synonymer
property
Collocations
- anvendelse af selskabets formue
- use of the company's assets
- beskatning af indkomst og formue
- taxation of income and property
- den ikke-erhvervsmæssige formue
- the non-business assets

(...)
Examples
- Fusion indebærer, at det selskab, der som en helhed overdrager sin samlede formue til et andet selskab, ophører med at eksistere, uden at der finder en egentlig likvidationsbehandling sted.

(...)
Synonymer
aktiver
kapital
værdier

activos (Spanish-English)

Definición

Los activos es el conjunto de bienes, propiedades y derechos propiedad de una persona física o jurídica.

assets

Definition

The totality of what a person or enterprise owns is called his or its assets.

Synonymer

property

Colocaciones

- el conjunto de los activos de una empresa
- all assets
- administrar activos
- administer assets

(...)

Ejemplos

- Los activos son un recurso o bien económico propiedad de un negocio, con el cual se obtienen beneficios. Los activos de un negocio varían de acuerdo con la naturaleza de la empresa.
- Assets are economic resources used for obtaining a profit. The assets of a company vary depending on the nature of the company.

Fuente

Wikipedia
[<http://es.wikipedia.org/wiki/Activo_(contabilidad][7]

The third implication is that users can retrieve items on demand, most typically by making use of different search options. For instance, in the *accounting dictionaries,* users can access 27 dictionaries, each of which allows them to search different data for the different use situation in which they find themselves. For example, a Spanish user in a reception situation will only retrieve the definition of, say, *assets* or *activos*, whereas they will retrieve synonyms, collocations, and examples in a production situation.

These reflections allow us to define multilingual lexicography in the era of the internet as the *theory and practice of unified and well-connected monolingual, bilingual and multilingual dictionaries using data from a multilingual database. These dictionaries are information tools that cover words, terms, facts, and/or things in several languages, have the same conceptualization at the pre-compilation phase, and make use of lexicographic and technological know-how which allows (a) lexicographers to add new languages to the same information database from which new monolingual, bilingual and multilingual dictionaries can be extracted and (b) users to retrieve connected data easily and to spot and understand possible similarities and differences among the several languages covered.*

[7] Last access: April 17, 2017.

4 Conclusion

This chapter situates the concept of multilingual lexicography in the era of the Internet in a new dimension. Firstly, it espouses the idea that multilingual lexicography should involve integrated information tools in which users, access and data are interconnected, based on the same conceptualisation, and approached holistically.

Secondly, it assumes that language is an umbrella term in lexicography. This means that multilingual dictionaries can and must cover words, terms, facts and things in a similar way, without establishing differences based on the ontological status of the lexicographic data in question.

Thirdly, it argues that multilingual dictionaries are flexible information tools. This has three implications: (a) they allow lexicographers to include new languages in the same information database from which new monolingual, bilingual and multilingual dictionaries can be extracted; (b) they also allow lexicographers to include specific features of each of the languages covered; (c) they allow users to employ different search systems and retrieve different data types, e.g. they can only retrieve monolingual, bilingual or multilingual data.

Finally, multilingual dictionaries must *also* offer users meaning items, especially one definition per sense, contextual clues, typically collocations and examples for each sense, and cultural information, e.g. lexicographic notes. To conclude, the multilingual dictionary in the era of the Internet is an *integrated* information tool that deals with several languages, offers complete descriptions of the lemmas covered, and uses up-to-date access systems that favour quick and easy consultation.

5 Bibliography

5.1 Dictionaries

5.1.1 Accounting Dictionaries

Nielsen, Sandro/Mourier, Lise/Bergenholtz, Henning/Almind, Richard, in collaboration with Grønborg, Helle/Melgaard, Mads/Middelboe, Trine/Sørensen, Brit (2003): Den Danske Regnskabsordbog. Center for Leksikografi: [<http://www.regnskabsordbogen.dk/iasdk>

Nielsen, Sandro/Mourier, Lise/Bergenholtz, Henning, with contributions by Johnsen, Mia/Bobjerg Nielsen, Rie/Kofoed Stender, Amalie/Vrang, Vibeke (2006): Den Engelsk-Danske Regnskabsordbog/English-Danish Dictionary of Accounting. Database og design: Richard Almind, implementering og hjemmesider: Caspar Thomsen: [<http://www.regnskabsordbogen.dk/regn/gbdk/gbdkregn.aspx>].

Fuertes-Olivera, Pedro A./Bergenholtz, Henning/Nielsen, Sandro/Gordo Gómez, Pablo/Mourier, Lise/Niño Amo, Marta/Ríos Rodicio, Ángel/Sastre Ruano, Ángeles/Tarp, Sven/Velasco Sacristán, Marisol (2012): Diccionario Inglés-Español de Contabilidad. Base de Datos y Diseño: Richard Almind and Jesper Skovgård Nielsen. Odense: [<Lemma.com>].

Fuertes-Olivera, Pedro A./Bergenholtz, Henning/Nielsen, Sandro/Gordo Gómez, Pablo/Niño Amo, Marta/Ríos Rodicio, Ángel/Sastre Ruano, Ángeles/Tarp, Sven/Velasco Sacristán, Marisol (2013): Diccionario Español de Contabilidad. Base de Datos y Diseño: Richard Almind y Jesper Skovgård Nielsen. Hamburg: [<Lemma.com>].

Diccionari de la Llengua Catalana Multilingüe: [<http://www.multilingue.cat/>].

Encyclopaedia Britannica: [<https://www.britannica.com>].

IATE = The EU's multilingual Term Base: [<http://iate.europa.eu/SearchByQueryLoad.do?method=load>].

Logos Dictionary: [<http://www.logos.it/>].

The New Palgrave Dictionary of Economics: [<http://www.dictionaryofeconomics. com/dictionary>].

5.2 Monographes and articles

Bergenholtz, Henning/Skovgård Nielsen, Jesper (2013): What is a Lexicographical Database? In: Lexikos 23, 77–87. [<http://lexikos.journals.ac.za/pub/article/ view/1205/716>; last access: April 17, 2017].

Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.) (2011): E-Lexicography: The Internet, Digital Initiatives and Lexicography. London/New York: Continuum.

Fuertes-Olivera, Pedro A./Tarp, Sven (2014): Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography. Berlin/Boston: De Gruyter.

Nielsen, Sandro/Almind, Richard (2011): From Data to Dictionary. In Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): E-Lexicography: The Internet, Digital Initiatives and Lexicography. London/New York: Continuum 141–167.

Tarp, Sven (2008): Lexicography in the Borderland Between Knowledge and Non-knowledge. Tübingen: Niemeyer.

Rufus H. Gouws

Metalexicographic models for multilingual online dictionaries in emerging e-societies

Abstract: The digital era poses numerous challenges to the planning and compilation of reference works, including dictionaries. Lexicographers need to respond to these challenges in a way that negotiates the needs and reference skills of their potential target users. This paper focuses on the South African lexicographic environment as an emerging e-society where the lexicographic needs but also the reference skills of a new generation of potential dictionary users need to be taken into consideration. Suggestions are made as to how new lexicographic products could assist the potential target users and how a more comprehensive dictionary culture can be established. An emphasis is placed on suggestions that smartphones should be seen as a preferred device for digital reference sources.

Keywords: bilingual dictionaries, e-dictionaries, dictionary culture, MobiLex, national lexicography unit, reference skills, smartphone, user needs

1 Introduction

In its development from the clay tablet to the internet, dictionaries and the field of lexicography have survived many transitions; not only with regard to the nature of the medium but also with regard to e.g. contents, structures, theoretical underpinning and even the ideological basis. Naturally, some of these changes have been more demanding and challenging than others but to a greater or lesser degree all of them had an influence that altered the course of the lexicographic practice at large. These transitions include the shift from the clay tablet and papyrus leaves to printed scripts, the shift from thematically ordered dictionaries to alphabetically ordered products, the shift from a prescriptive to a descriptive approach, the shift from a card collection to an electronic corpus, the shift from a random theory-detached to a theoretically-based approach and, significant for this paper, the shift from printed to online dictionaries.

All these different transitions had an impact on dictionaries but also on both lexicographers, and dictionary users. In this regard the following section will refer to only a few watershed moments in the development of the lexicographic practice – primarily using examples from the English-speaking lexicographic environment.

Rufus H. Gouws: Department of Afrikaans and Dutch, University of Stellenbosch, Private Bag X1, Matieland, 7602 South Africa, tel. +27 218082164, fax +27 218083815.

https://doi.org/10.1515/9783110607659-003

This will be followed by a focus on the influence and expectations of the online transition in multilingual and multicultural societies where the internet has not yet been established as the default medium for information retrieval, with the South African[1] lexicographic environment being the case in point.

2 Significant changes

The early dictionaries compiled on clay tablets in the then Persia when the Assyrians invaded Babylon, consisted primarily of a list of Assyrian words with their Sumerian equivalents, see McArthur (1986). A significant feature of these pioneering dictionaries was their user-directed aim. They were compiled as practical instruments in response to the needs of real people with real needs and these dictionaries had to satisfy the needs that these people experienced in real life situations. Centuries later with a well-established tradition of thematically ordered dictionaries where lexicographers wanted to put the continuum of life into manageable containers by ordering their *vocabularia* in themes and topics to reflect something of the world order, Robert Cawdrey's TA (1604) brought a wholly new emphasis. He stated clearly in the preface that the intention with this dictionary was to give an explanation of the meaning of difficult words, gathered for the help of ladies, gentlewomen or any other unskillful person. The dictionary was yet again seen as a practical instrument to assist ordinary people in their quest for help.

While working on his DEA that was published in 1755 Samuel Johnson's idea was that the dictionary should be an instrument to fix the language, seeing that he was distressed by the low standard of the prevailing English usage. This led to a strongly prescriptive approach by many lexicographers –an approach that still plays a major role in many dictionaries of our time.

In 1961 Philip Gove's W3 was published and few if any dictionaries before or after this dictionary has caused such a public outcry from linguists, dictionary users and the public at large, cf. in this regard Sledd/Ebbitt (1962). The extremely negative criticism was mainly due to Gove's approach to include actual language and not only linguistically pure or proper words. His argument was that a dictionary has to reflect language, not set its style. Another paradigm shift in the lexicographic practice was the publication of John Sinclair's COBUILD (1987). This dictionary introduced the use of an electronic corpus displaying actual current language use,

[1] This research is supported in part by the National Research Foundation of South Africa (grant specific unique reference number (UID) 85434). The Grantholder acknowledges that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the author, and that the NRF accepts no liability whatsoever in this regard.

focused on the needs of learners with the promise, given on the cover page, of "helping learners with real English."

In all these changes the medium of dictionaries remained the same, in a broad sense the printed medium whether produced on clay tablet, papyrus leaves, parchment or paper. The most dramatic change in recent times, that can probably also be regarded as the most significant throughout the history of dictionary-making, has been the transition from printed to online dictionaries. This is not only a transition from one medium to another but at the same time also a watershed in terms of presentation, access and contents, characterized by a vast array of new possibilities, opportunities and challenges. In this regard Rundell (2012: 72) says:

> The migration from print to digital is the second big upheaval for lexicography in the last 30 years. The corpus revolution forced a major rethink of lexicographic practice in both 'analysis' and 'synthesis' modes (as well as changing our perceptions of how language works). Yet the changes it led to have been mainly 'internal', affecting the way lexicographers work and improving the reliability of their output. The end-product is still recognizably a dictionary, and for the average user the changes going on behind the scenes may be barely perceptible. But the new and ongoing digital revolution will be more disruptive. Its effects are 'external', in that it impacts directly on dictionary users, and is in a sense driven by their changing behavior.

There has been a less dramatic transitional phase with the introduction of dictionaries on CD-ROM; with the most important contribution being the indication of some possibilities of a new medium awaiting lexicographers. Although these dictionaries on CD-ROM were a vast improvement in terms of access and certain search procedures they remained in the most instances mere digital versions of printed dictionaries. This type of e-dictionary will not fall within the scope of the discussion in this paper.

Although the transitions mentioned in this section belong to the past, current and future lexicographers could do well to reconsider some of these issues in the planning and compilation of online dictionaries.

3 A new era

The transition from the printed to the online era is a multi-phased and time consuming process which currently is well underway but not yet completed. It is a process with a staggered start –both regarding different communities embarking on the process and individuals or groups within a single community. All geographical and social regions did not and could not embark on this process simultaneously or commit themselves to the same extent.

The implications of the internet and the online possibilities and challenges for lexicography are not only relevant for the lexicographic practice but also for metalexicography and the assessment of and adaptations to existing lexicographic theories. Prevailing lexicographic theories have primarily been devised in the era of printed dictionaries and although all theories and all components of these theories are not medium-specific and therefore not restricted in their application to the planning and compilation of printed dictionaries metalexicographers need to negotiate the adaptations that need to be made to provide for the demands of the online lexicographic practice.

The sphere and scope of online lexicography should not be seen as restricted to the lexicographic practice. Lexicographic theory stands at the heart of this new phase in lexicography. Important in this regard is the fact that lexicographers need to be careful not to repeat the mistakes made in the era of printed dictionaries and the accompanying theoretical discussions, cf. Gouws (2011). Online dictionaries should not be products that lack a theoretical basis and theoretical lexicographers should hasten to formulate the necessary models to ensure the successful planning and compilation of online dictionaries.

The application of a sound lexicographic theory in the practice of online lexicography should ensure the acknowledgement of various types of diversity that need to be negotiated. This includes the nature and extent of the prevailing dictionary culture –or the lack thereof. In this regard it is especially the societal dictionary culture and not the ideolectal dictionary culture, cf. Gouws (2012) that should play a determining role in the planning of online dictionaries. Linguistic and cultural diversity confront lexicographers with a wide range of challenges –this applies to printed and online dictionaries. Yet again, innovative planning is needed to ensure the best possible outcome and not a mere repetition of insights from the past. Diversity also applies on the technological level –with regard to the software, planning and compilation of online dictionaries but also with regard to the technical expertise of the potential target users of the envisaged dictionary in using e-devices. In the planning of any dictionary the lexicographer should not only identify the target user but should also be aware of the needs and the reference skills of that user. These reference skills should play an important role in the decisions regarding e.g. the presentation of data and the structures of a given dictionary. Online dictionaries demand their own reference skills and in the planning of new dictionaries lexicographers need to take cognizance of the skills of their envisaged users because this could have an influence on the nature of the dictionary to be compiled. In the planning of online dictionaries lexicographers should not take the level of skills their users have with regard to printed dictionaries as a criterion to determine their potential efficiency and skills with regard to online dictionaries. Different generations have different skills in different domains and media and lexicographers need to be aware of this fact.

The online era has not yet come to fruition in all societies. Unfortunately, where it has been established lexicographers often did not learn from the transitions of the past in their planning and compilation of online dictionaries. Embracing some of the seducing novelties of the online environment they have compiled dictionaries that lack a theoretical basis and a clearly identified target user group and they have failed to negotiate the needs and reference skills of the potential dictionary users. On the bright side, where the online era has not yet come of age lexicographers still have the opportunity to plan before they compile their dictionaries and to adhere to guidance from theoretically-based lexicographic models. In societies where the internet is not yet the default reference source lexicographers can now plan way in advance so that they can apply these models as soon as it is technically possible and feasible in their societies.

Lexicographers should always be aware of the need for a good relation between lexicography and society. Lexicography has a very specific responsibility towards society and this has to be negotiated in a new way for e-lexicography. User-friendliness in online lexicography may demand different approaches, e.g. different types of dictionaries and different types of treatment compared to user-friendliness in the domain of printed dictionaries. A more comprehensive dictionary culture is needed that embraces a bidirectional relation between lexicography and society, cf. Gouws (2016), and that makes provision for all aspects of lexicography, including the medium of dictionaries and the society for which they are compiled. This paper will focus on some aspects of such a dictionary culture, i.e. the process of embarking on online lexicography in a multilingual and multicultural society where the internet has not yet had a full impact in all spheres of life, a society to be characterized as an emerging e-society.

4 Emerging e-societies

In this paper the term *emerging e-societies* is used to refer to societies where internet access is not yet fully available and where online sources are not yet regarded as the default source to be used for the retrieval of information. In the field of lexicography it would also imply that printed dictionaries still play a dominant role in the reference endeavors of some people and online dictionaries still need to cross the threshold of high frequency use. The transfer from printed to online dictionaries in these societies should be assessed with cognizance to the prevailing dictionary and reference culture. Many African speech communities, including some in South Africa have an insufficient dictionary culture, not only with regard to online dictionaries but also still with regard to printed dictionaries and the insufficient dictionary culture or even the total lack thereof is one of the results of a lack of available printed dictionaries or a lack of access to these dictionaries. Lexicographically some of these societies still have a Middle Ages experience –with dictionaries not even available for ladies, gentlewomen or any other unskillful person!

Various projects, e.g. in South Africa by the National Lexicography Units of PanSALB, i.e. the Pan South African Language Board, are currently underway to supply printed dictionaries to all speech communities of the eleven official South African languages. This supply of dictionaries is unfortunately not sufficiently complemented by attempts to empower children and other potential users of these dictionaries with the necessary dictionary using skills. The role of dictionaries as practical instruments that assist their users in the process of life-long learning has unfortunately not yet been grasped by the majority of stakeholders in the different educational processes. This applies to printed dictionaries but in the planning of online dictionaries for these societies efforts should be made to prevent a similar situation.

The lack of a dictionary culture and the non-availability of dictionaries have resulted in a number of lexicographically lost generations in South Africa. This poses an interesting challenge to South African lexicographers. The question is whether attempts should be made to introduce a printed dictionary-based dictionary culture as an obligatory first phase in the dynamic process of acquiring a comprehensive dictionary culture or whether this phase could be omitted in favor of an online dictionary-based approach as the first phase. Should this lexicographic Renaissance follow the Middle Ages or the Post-Modern era? The answer to this question also demands input from the field of education and, specifically, the didactics of e.g. language learning, information science and knowledge retrieval. But it also demands common sense and an acknowledgement of the reality and the digital skills of younger generations, even in an emerging e-society.

Teaching someone to drive a horse cart will not help him/her to become a better driver of a motor car. Likewise, compelling someone to learn to use a printed dictionary will not necessarily increase their ability to use an online dictionary. Consequently, one line of thought, adhered to in this contribution, is to look forward and to plan accordingly; forgetting about prevailing lexicographic problems and insufficiencies. However, this planning may not be done in a way where theory follows the practice. Theory should point the way so that the lexicographic practice can succeed. The young and next generations of dictionary users, also in developing countries and in emerging e-societies, will eventually have access to the internet and will be skilled and trained in retrieving the information needed for their daily life from this source –even though they may never have had access to a printed dictionary.

In order to attempt to have dictionaries regarded and used by these speakers as practical tools, lexicographers need to ensure that they have access to dictionaries in their preferred medium of information retrieval. Even though online access is still only available in a restricted manner the young generation should be introduced into the world of online dictionaries. Not forgetting the problems due to insufficient access to printed dictionaries and the subsequent lack of dictionary using skills, this paper will focus on ways to go about in ensuring a better lexicographic future by means of online dictionaries.

Within the framework of the DAAD call "Welcome to Africa" SeLA, i.e. Scientific e-Lexicography for Africa, a joint project between universities in Germany, South

Africa and Namibia, is currently in its final phase. This project has focused on both teaching and research to enhance the planning of and access to e-lexicography products for the South African environment. The idea has not been to compile new dictionaries but to assess the situation, analyze the user needs and devise new ways for user interaction, supply the required theoretical guidance and apply it to prepare some mock-up dictionaries that could enable improved access to relevant data.

From the work done within the SeLA project but also from experience of the South African lexicographic environment it is clear that the challenges facing lexicographers need to be dealt with in a situation-specific way.

Given the situation in emerging e-societies one has to provide assistance that could contribute as good as possible to solving the diverse lexicographic needs of the intended users. Albeit that the needs will mostly fall within the domain of communication functions the multicultural environment should also be negotiated. Dictionaries should therefore assist their users in bridging both linguistic and cultural gaps and should satisfy both communication and cognitive functions. Yet again, all these challenges will not be met simultaneously in the first lexicographic product to be made available, but the problems need to be addressed in subsequent phases of a comprehensive lexicographic process.

In the remainder of this paper the focus will be on finding possible solutions for various problems prevailing in the South African lexicographic situation. The proposals to be made in this discussion are not only relevant for South Africa but could be adapted and applied in the lexicographic endeavors in other emerging online societies.

5 The language and lexicographic situation in South Africa

Although South Africa has eleven official languages they are not all spoken throughout the country and the use of some of these languages are subjected to stringent geographical restrictions. English and Afrikaans have the most widespread use, with English, a language of documentation, to be regarded as the lingua franca. The lexicographic development of these languages differs considerably, with some languages, like Afrikaans and South African English, having a variety of dictionaries responding to a wide spectrum of user needs, whereas speech communities of some of the other languages have little lexicographic assistance in their own languages.

Due to its multilingual and multicultural nature South Africa can be regarded as a lexicographic laboratory that lends itself to different experiments with regard to the planning and compilation of dictionaries, research regarding dictionary use and problems in the transition from printed to online dictionaries. The opportunities as well as the need for lexicographical intervention are obvious and different responses to this situation are required. Among others there is a real need for lexicographers to embark on new projects that respond to the diverse lexicographic and reference needs of the different speech communities and for publishing houses to make the required dictionaries available –in the desired format and medium with the desired functions and content.

In principle one can be optimistic about the future of South African lexicography. Besides the important and comprehensive lexicographic work done by commercial publishers, albeit not equally directed at all the languages, South Africa is in the potentially favorable position of having a state-funded national lexicography unit (NLU) for each of the official languages, a situation to which Gouws (2000: 114) refers as "a golden opportunity". The brief of these units is to provide in the lexicographic needs of the speech communities they represent. This is only a potentially favorable situation because the practice has unfortunately not shown sufficient evidence of significant progress since the establishment of these NLUs in 2001. The productivity of some of the NLUs as well as the quality of their products can be questioned. These problems will not be discussed in this paper but one has to bear the potential contribution of those NLUs that do not perform optimally in mind when planning a comprehensive approach where collaboration between NLUs is needed. Currently the NLUs operate in isolation without real cognizance of the work done in the other units and their efforts are mostly directed at the production of printed dictionaries –both monolingual and bilingual dictionaries. In the bilingual dictionaries the language of the specific NLU is mostly paired with English. The exceptions to this approach are the NLUs for English and Afrikaans. The unit for Afrikaans is busy with a longstanding project, i.e. a comprehensive multivolume monolingual dictionary, whereas the unit for English is involved in various monolingual dictionaries of South African English.

Looking at online dictionaries in South Africa the results are not as exciting as one would have expected. The WAT, currently the comprehensive project of the NLU for Afrikaans, was started in 1926 and 14 volumes, covering the alphabet from A-Skooi, have already been published. The completed article stretches are also accessible online. Some products from other NLUs, e.g. the DSAE, compiled by the NLU for English, are also available online. However, these dictionaries as well as a vast number of dictionaries from commercial publishers that are available online have not initially been planned as online dictionaries. The online products are mere digitized versions of the printed dictionaries. Although the South African languages host some established lexicographic work the planning of new projects and products have not been directed sufficiently at the online medium and have failed to utilize the new possibilities.

Martin (1996) already gave some indications of a digitized lexicographic process for the South African environment. Martin (1996: 204) acknowledges the fact that a multilingual environment may require a special kind of translation dictionary and in this regard he proposed a multifunctional dictionary, serving the needs of both source and target language speakers. Referring to the proposals made by Mashamaite (1995), Martin (1996: 209) proposed his well-known hub-and-spoke model that could provide for several bilingual dictionaries where lexical items of the spoke languages are linked to a common hub language. Within the South African environment this will imply that English, as lingua franca, will be the hub language. Lexical items from the ten remaining official languages will be linked to English. A more advanced application of the model could also provide the linking of spoke languages to one another.

Utilizing the advantages of major developments in the field of online lexicography since the suggestions by Martin (1996) the idea of a multilingual online dictionary in which the eleven official languages of South Africa are linked still seem to be a deserving idea. The planning of such a project needs to go way beyond the production of only a single multilingual dictionary. Much more attractive is the idea of a comprehensive database from which different dictionaries can be extracted that respond to the needs and profiles of specific users or user groups. The eventual aim of such a process could be to have a South African dictionary portal that accommodates monolingual, bilingual and multilingual dictionaries for and between all the official South African languages. The NLUs are in the ideal position to play a leading role in such a process.

The online era should have an influence on the work of the different NLUs and their efforts also need to be guided towards producing online dictionaries. Besides continuing with some of their current projects the NLUs could also combine forces and expertise and embark on a bigger lexicographic endeavor to ensure the beginning of a South African multilingual online dictionary project. The first phase of such a project could be directed primarily at the young members of the e-generation –those who have been born in and into an online world. Albeit that they may not yet have sufficient access to the internet using computers or tablets as instruments, these users are well familiar with smartphones. Du Plessis (2014: 78) indicates that 40% of the population of South Africa are smartphone users and school children represent a large section of these users. Consequently a first joint attempt by the NLUs could be to devise a model that will give these users access to online dictionaries via their smartphones. Such a project can be of immediate and real importance.

6 The subject matter of the envisaged dictionaries

When planning online dictionaries for the South African situation it is important to decide at an early stage of the process what the subject matter and the broad typological categories of the envisaged dictionaries should be. The history of printed dictionaries shows that bilingual dictionaries are often the first lexicographic products made available to speakers of non-standardized languages, cf. Gallardo (1980). Bilingual dictionaries also play a meaningful role in multilingual societies and represent a frequently used dictionary type. Printed bilingual dictionaries have occupied this position but this important role should also be considered in the planning of the early representatives of online lexicography.

Phasing a society into online lexicography and phasing online dictionaries into a society should be done in a careful but scientifically-based way. One should work with the assumption that an online dictionary for an emerging e-society will differ from a comparable dictionary for an established e-society. Albeit that the members of such a society have varying lexicographic needs all these needs could not be satisfied in the first dictionary. Lexicographers could decide to focus on a product to satisfy a first level of needs of an emerging group of dictionary users instead of focusing on the needs of the majority of potential target users. Collaboration between lexicographers from the different NLUs could lead to the planning and compilation of a single multilingual online dictionary as valuable reference tool for the envisaged target users. Although the compilation of this online dictionary may also proceed in staggered phases the subject matter will eventually be all the different official languages of South Africa.

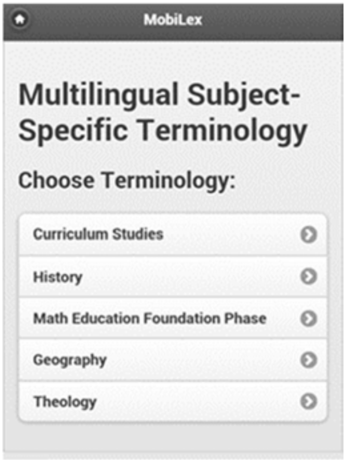
Using a single database with a lemma selection provided by the NLU for South African English the lexicographers from the other units can provide equivalents for these English lemmata, leading to ten bilingual dictionaries with English as one member of the language pair. Subsequent processes may then lead to bidirectional bilingual dictionaries, to more advanced versions of these dictionaries that present the basic vocabulary of each language, to a linking of the spoke languages to ensure bilingual dictionaries with language pairs that do not include English as a treated language and to the eventual production of monolingual dictionaries for each of the languages. All these variations could be compiled from the same database.

7 Models for the online generation

The lexicographic skills with regard to printed dictionaries of the e-generation may be lacking but the same does not apply to their computational and especially their cell phone skills. The smartphone could be the best possible instrument to target as a way of introducing young members of the e-generation to lexicography. In these attempts it is important to work with the assumption that the traditional notion of a dictionary may have become obsolete for the specific target user group. Their real needs are still predominantly access to data and therefore to the most appropriate information tool. Consequently it is important that the planning of a database should not only be directed at the retrieval of information from the data presented in specific dictionaries but also at giving access to data that fall beyond the scope of traditional lexicographic products and therefore also beyond the dictionaries included in the database. The planning of these dictionaries and such a multilingual database should make provision for the inclusion of data relevant to satisfying both communication and cognitive needs. All these needs will not necessarily be dealt with in all the different dictionaries but the database can include additional data that can be accessed when needed.

In this paper the technical aspects regarding the smartphones, database and software relevant to the proposals will not be discussed. The focus will rather be on a metalexicographic wish list.

At Stellenbosch University colleagues from the Department of Curriculum Studies in the Faculty of Education are working on a pilot project, MobiLex, a web-mobile application that allows users to search for terms from certain subject fields, i.e. currently the fields of curriculum studies, history, mathematical education for the foundation phase, geography. The Faculty of Theology has added the broad subject field of theology for their students to this list. When accessing MobiLex the user has to choose a given subject field, cf. Figure 1:


MobiLex

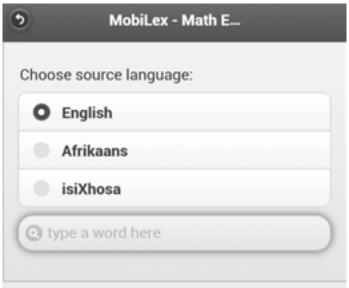
Multilingual Subject-Specific Terminology

Choose Terminology:

- Curriculum Studies
- History
- Math Education Foundation Phase
- Geography
- Theology

Fig. 1: MobiLex

The user then has the option to enter the required source language, choosing between Afrikaans, English and Xhosa, cf. Figure 2:



MobiLex - Math E...

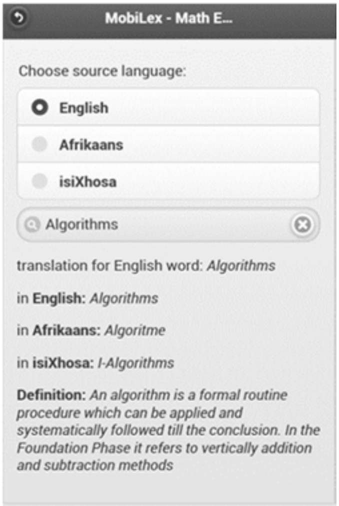
Choose source language:

● English
○ Afrikaans
○ isiXhosa

type a word here

Fig. 2: Selecting the source language in *MobiLex*

A next step requires the user to enter a term from the specific subject field. This leads to a screen shot that displays the English, Afrikaans and isiXhosa equivalents as well as a brief explanation of meaning, giving in the language entered as source language. Taking cognizance of the restricted screen space on a smartphone the explanation of meaning is limited to 250 characters, cf. Van der Merwe (2015). The article presenting the treatment of such a term is seen in Figure 3:



MobiLex - Math E...

Choose source language:

● English
○ Afrikaans
○ isiXhosa

Algorithms

translation for English word: *Algorithms*

in **English:** *Algorithms*

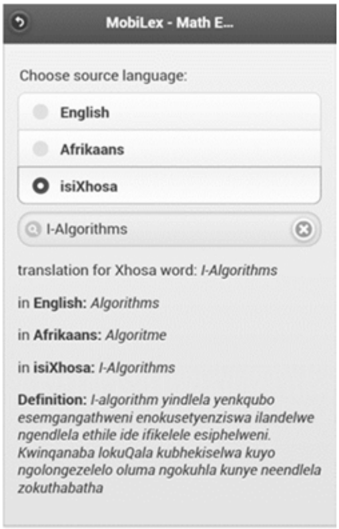
in **Afrikaans:** *Algoritme*

in **isiXhosa:** *I-Algorithms*

Definition: *An algorithm is a formal routine procedure which can be applied and systematically followed till the conclusion. In the Foundation Phase it refers to vertically addition and subtraction methods*

Fig. 3: Selecting the equivalents in *MobiLex*

A mother-tongue speaker of Xhosa who is not familiar with the English term *algorithm* can find the Xhosa translation equivalent and then, with Xhosa as source language, enter the Xhosa equivalent to find the explanation of meaning in this language, cf. Figure 4:



Choose source language:

○ English

○ Afrikaans

● isiXhosa

I-Algorithms

translation for Xhosa word: *I-Algorithms*

in **English:** *Algorithms*

in **Afrikaans:** *Algoritme*

in **isiXhosa:** *I-Algorithms*

Definition: *I-algorithm yindlela yenkqubo esemgangathweni enokusetyenziswa ilandelwe ngendlela ethile ide ifikelele esiphelweni. Kwinqanaba lokuQala kubhekiselwa kuyo ngolongezelelo oluma ngokuhla kunye neendlela zokuthabatha*

Fig. 4: Sample caption

Although MobiLex displays a very basic lexicographic approach it is a significant first step to familiarize users with the concept of finding lexicographic assistance on their smartphones. It serves restricted text reception purposes and, albeit inadequately, also minimal translation purposes, although no contextual guidance is given to ensure the proper selection and correct use of a translation equivalent.

MobiLex, however, is not the multilingual dictionary envisaged for the NLUs, but a similar approach can be used in the joint project, focusing eventually on both terms from different subject fields and lexical items from the language for general purposes. Comparable to MobiLex a first version of such a dictionary could present a brief lexicographic presentation and treatment of words and subject-specific terms from all the different South African languages. If this envisaged project can introduce a dictionary on a smartphone in the early school years that can really assist the users successfully on a daily basis inside and outside the classroom the users could be convinced of the value of a dictionary as information retrieval instrument.

In order to stimulate potential users to use such a dictionary its subject matter and the specific treatment allocated to lemmata should go beyond the typical language and even lexicographic domain. Gouws (2012) already argued in favor of a so-called system of integrated dictionary use. This implies among others the planning of a school dictionary in such a way that it can be linked to text books and other work done in a specific curriculum. This approach could also be employed in the planning of smartphone dictionaries for the early (and even later) school years. A dedicated corpus could be developed for the different school subjects relevant to a given user group. Such a corpus could contain subcorpora for different age groups.

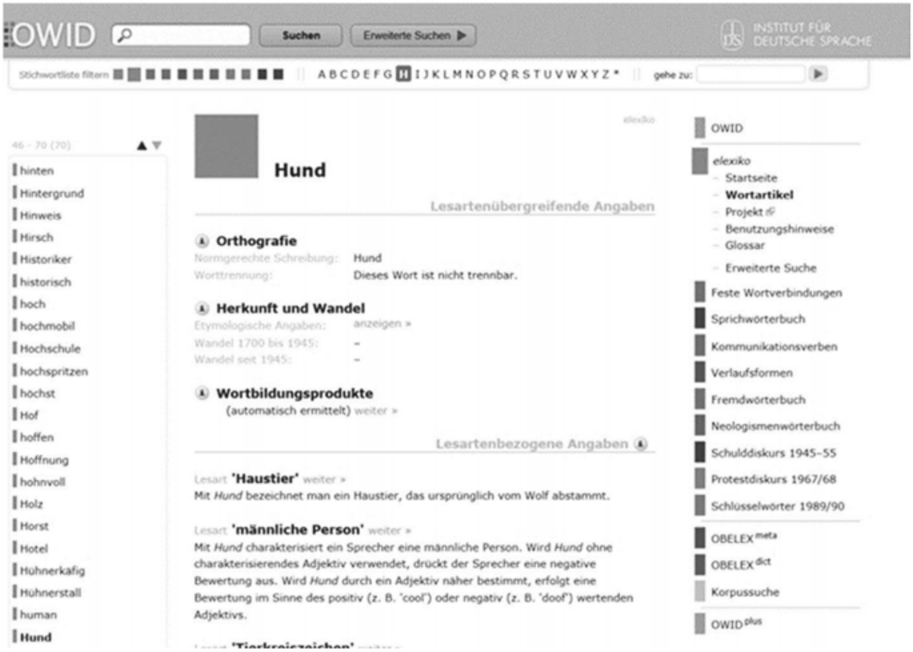
In this regard it is important to have the cooperation of all the NLUs –each unit supplying the data for their respective language from their respective corpora. Such a project relies on thorough planning and agreement on the genuine purpose of the envisaged dictionary series. If the project starts with multilingual dictionaries that have their focus on terms from the school curricula these terms may initially still be isolated from the text books, but if these text books can also be included in the database the user can move between the treatment in the text book and the online dictionary. Using a smartphone to move between these different sources could form the basis of further reference possibilities. The lexicographers participating in this project need to take cognizance of not only the lexicographic needs of their potential target users but also their general reference needs. Where the lexicographic treatment could be the first layer of data from which information can be retrieved, the dictionary articles could also include links not only to text books but also to other references, including extra-curricular sources and even search engines like Google. This could help to ascertain and enhance the role of dictionaries as more general information tools and to give them a firm footing within the e-generation.

Where MobiLex currently focuses in a restricted way on communicative functions a dictionary that has the needs of both an intra-school and an extra-school living environment in its scope should probably also focus on a cognitive function. Besides access to extra-dictionary sources that could enhance the cognitive function the cognitive function should also be strengthened on an intra-dictionary level. This demands the inclusion of some additional data in the dictionary articles. One aspect to be considered in a multilingual and multicultural environment where dictionaries need to fulfil a cognitive function is the occurrence in all the South African languages of culture bound lexical items. These items are used in the different languages but often also included as loan words in the other South African languages – where the speakers are not familiar with the specific cultural concepts. In this regard lexicographers from the NLUs should make a selection of these items from each language. Their inclusion and treatment could enhance the text reception assignment of a dictionary but could also support the cognitive function and will assist the dictionary users in their daily exposure to different languages and cultures.

When consulting online dictionaries users are too often confronted by an information overload. This is a predictable result of the misconception that e-dictionaries have no space restrictions, cf. Rundell (2012: 73). Smartphones have even more stringent space restrictions than many of their printed counterparts. These restrictions apply to what can be presented in a single screenshot. Lew (2012) distinguishes between the infinite *storage space* and the restricted *presentation space,* i.e. the space on a computer screen. The latter determines how much can be presented at any given time to the user. The presentation space should play a determining role in the planning of smartphone dictionaries.

More space freedom does prevail when employing a system where a single dictionary article has a layered presentation that utilizes different screenshots –

without trying to clutter too many entries into a single screenshot. In this regard Gouws (2014) discusses aspects of the choices a dictionary like *elexiko* gives its users to access those search zones in an article from which they need to retrieve information –without having to read through the whole article. A smartphone could have a similar approach where the opening screenshot can display the menu on offer for a word entered as search word. The user then has the option to move to the screenshot that will display the data he/she might need. For the word *Hund* ('dog') *elexiko* has the following screenshot that displays different search zones, including different subcoments on semantics, and the user has the option to click on the given links to access additional data:



Hund

Lesartenübergreifende Angaben

Orthografie

Normgerechte Schreibung: Hund
Worttrennung: Dieses Wort ist nicht trennbar.

Herkunft und Wandel

Etymologische Angaben: anzeigen »
Wandel 1700 bis 1945: –
Wandel seit 1945: –

Wortbildungsprodukte

(automatisch ermittelt) weiter »

Lesartenbezogene Angaben

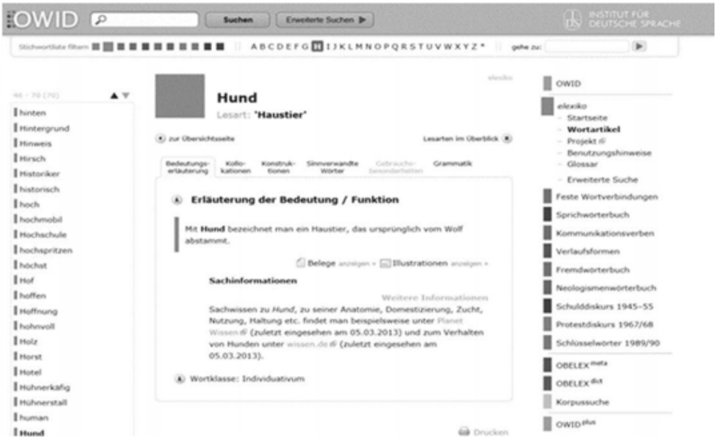
Lesart **'Haustier'** weiter »
Mit *Hund* bezeichnet man ein Haustier, das ursprünglich vom Wolf abstammt.

Lesart **'männliche Person'** weiter »
Mit *Hund* charakterisiert ein Sprecher eine männliche Person. Wird *Hund* ohne charakterisierendes Adjektiv verwendet, drückt der Sprecher eine negative Bewertung aus. Wird *Hund* durch ein Adjektiv näher bestimmt, erfolgt eine Bewertung im Sinne des positiv (z. B. 'cool') oder negativ (z. B. 'doof') wertenden Adjektivs.

Lesart **'Tierkreiszeichen'** weiter »

Fig. 5: An article struture in *elexiko*

By clicking on the link "weiter" ('further') in the first subcomment on semantics "Haustier" ('domestic animal') the user is guided to the following screenshot:



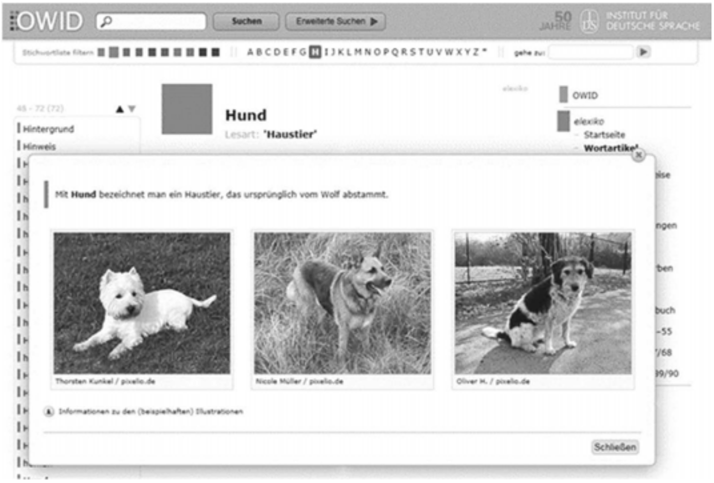
Mit **Hund** bezeichnet man ein Haustier, das ursprünglich vom Wolf abstammt.

Sachwissen zu *Hund*, zu seiner Anatomie, Domestizierung, Zucht, Nutzung, Haltung etc. findet man beispielsweise unter Planet Wissen (zuletzt eingesehen am 05.03.2013) und zum Verhalten von Hunden unter wissen.de (zuletzt eingesehen am 05.03.2013).

Fig. 6: A subcomment on semantics in *elexiko*

A click on the link "anzeigen" ('show') next to the word "Illustrationen" ('illustrations') guides the user to the screenshot in figure 7.

The planning of the smartphone dictionary could make provision for comparable choices according to the needs of the target users. Where a user is familiar with the specific dictionary and usually consults it for a specific type of assistance, e.g. items giving the translation equivalents or items giving the paraphrase of meaning, the specific screenshot can be set as default opening screen that is displayed after a search word has been entered.



Mit **Hund** bezeichnet man ein Haustier, das ursprünglich vom Wolf abstammt.

Fig. 7: Pictorial illustrations in *elexiko*

8 Conclusion

Lexicography has a definite obligation and responsibility towards society. In the digital era this responsibility may not be diminished or underestimated. Lexicographers should realize that this responsibility can even go beyond the scope provided for by traditional dictionaries. Dictionaries need to be planned as information tools that offer access to both traditional lexicographic data but also to selected extra-lexicographic data. In emerging e-societies with a prevailing lack of a dictionary culture lexicographers should embark on a process to establish a comprehensive dictionary culture that includes a focus on the use of both online and printed dictionaries but where the online dictionary culture can exist without a prior dictionary culture directed at printed dictionaries. Attempts should not in the first instance be made to save lost lexicographic generations but rather to prevent further lost generations. The young members of emerging e-societies will usually be knowledgeable users of smartphones before they have access to other e-devices like computers and tablets. Consequently lexicographers should endeavor to use smartphones as instruments that offer them their first introduction to dictionaries and the world of reference.

9 Bibliography

9.1 Dictionaries

9.1.1 Printed dictionaries

COBUILD = Collins COBUILD English Language Dictionary. Ed. by Sinclair, John. London: Collins, 1987.

DEA = A Dictionary of the English Language. Johnson, Samuel. London: J./P. Knapton, T./T. Longman et al., 1755.

DSAE = A Dictionary of South African English on Historical Principles. Ed. by Silva, Penny. Cape Town: Oxford University Press, 1996.

TA = A Table Alphabeticall. Cawdrey, Robert. London: E. Weaver, 1604.

W3 = Webster's Third New International Dictionary of the English Language. Ed. by Gove, Philip, B. Springfield, Massachusetts: Merriam-Webster, 1961.

WAT = Woordeboek van die Afrikaanse Taal. Ed. by Botha, Willem F. et al. Stellenbosch: Buro van die WAT, 1951.

9.1.2 Online dictionaries

DSAE = A Dictionary of South African English on Historical Principles. [<http://dsae.co.za/>].
Elexiko = [<http://www.owid.de/wb/elexiko/start.html>].
Mobilex = [<http://www0.sun.ac.za/mobilex/>].
WAT = Woordeboek van die Afrikaanse Taal. [<http://www.woordeboek.co.za.ez.sun.ac.za/>].

9.2 Monographes and articles

Du Plessis, André (2014): A Functional Analysis of the e-WAT with Specific Focus on the Mobile Version: Towards a Model for Improvement. In: Lexikos 24, 75–93.

Gallardo, Andrés (1980): Dictionaries and the Standardization Process. In: Zgusta, Ladislav (ed.): Theory and Method in Lexicography. Columbia: Hornbeam Press, 59–69.

Gouws, Rufus H. (2000): Toward the Formulation of a Metalexicographic Founded Model For National Lexicography Units in South Africa. In: Wiegand, Herbert E. (ed.): Wörterbücher in der Diskussion IV. Tübingen: Max Niemeyer, 109–133.

Gouws, Rufus H. (2011): Learning, unlearning and innovation in the planning of electronic dictionaries. In: Fuertes-Oliviera, Pedro A./Bergenholtz, Henning (eds.): e-lexicography. London: Continuum, 17–29.

Gouws, Rufus H. (2012): Towards a system of integrated dictionary use. In: Karpova, Olga/ Kartashkova, Faina (eds): Multi-disciplinary Lexicography: Traditions and Challenges of the XXIst century. Cambridge: Cambridge Scholars Publishing, 134–144.

Gouws, Rufus H. (2014): Article Structures: Moving from Printed to e-Dictionaries. In: Lexikos 24, 155–177.

Gouws, Rufus H. (2016): Op pad na ʼn omvattende woordeboekkultuur in die digitale era. In: Lexikos 26, 103–123.

Lew, Robert (2012): How can we make electronic dictionaries more effective? In: Granger, Sylviane/Paquot, Magali (eds.): Electronic Lexicography. Oxford: Oxford University Press, 343–361.

Martin, Willy (1996): Lexicographical Resources in a Multilingual Environment: An Orientation. In: Lexikos 6, 199–214.

Mashamaite, K.J. (1995): The Hub-and-spoke Model: A Recipe for Making Bilingual Dictionaries between African Languages in South-Africa. Doctoral dissertation. Free University of Amsterdam.

McArthur, Tom (1986): Worlds of Reference. Cambridge: Cambridge University Press.

Rundell, Michael (2012): It works in practice but will it work in theory?' The uneasy relationship between lexicography and matters theoretical. In: Vatvedt Fjeld, R./Torjusen, Matilde J. (eds.): Proceedings of the 15th EURALEX International Congress. 7-11 August 2012. Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo, 47–92.

Sledd, James/Ebbitt, Wilma R. (eds.) (1962): Dictionaries and That dictionary. Chicago: Scott, Foresman and Co.

Van der Merwe, Michele (2015): Perceptions of the MAAL generation in higher education on the use of LSP dictionaries on mobile phones. Paper presented at the StelleLex-Colloquium, Stellenbosch, March 2015.

Sven Tarp

A dangerous cocktail: databases, information techniques and lack of vision

Abstract: This contribution discusses challenges to lexicography created by the new computer, information and communication technologies and techniques. It argues that the current transition period is full of paradoxes and that the main problem seems to be the subjective factor, i.e. the ability to adapt fully to the new technologies and get rid of old habits and ways of thinking. The article provides some examples of how the current challenges can be approached in terms of databases, user interfaces and other tools and techniques to assist the compilation and presentation of online dictionaries. In this context, it also calls for the opening of new areas of research into the empirical basis of dictionary production. The contribution defends the need for a lexicographical theory and a theory-based methodology which should be combined with visions on how lexicography and technology can be integrated in an ever ascending spiral that constantly provides new solutions to both old and new problems.

Keywords: e-lexicography, lexicographical databases, lexicographical user interfaces, information techniques, lexicographical alienation, function theory

1 Introduction

It is becoming trivial to state that lexicography is passing through a profound transition from the printed to the digital media. It is, however, anything but trivial to discuss and analyse the roots of this transition as well as its many challenges.[1]

If we briefly review the cognitive history of human beings from our origin as a species to the present days, we will see how our knowledge –and skills– have developed in the dialectics between contentedness with the old and well-known and curiosity with the new and unknown, between fear and audacity, routine and experiment, imitation and creation. Within specific areas of activity, this dialectics expresses itself in relatively long periods of quantitative accumulation followed by

[1] Thanks are due to the Spanish Ministry of Economy and Competitiveness for funding the project "La Teoría Funcional de la Lexicografía: Diseño y Construcción de Diccionarios de Internet" (Ref. FFI2014-52462-P).

Sven Tarp: Aarhus University (Denmark), University of Valladolid (Spain), Stellenbosch University (South Africa) and Guangdong University of Finance (China), Jens Chr. Skous Vej 4, 8000 Aarhus C (Denmark), st@cc.au.dk

https://doi.org/10.1515/9783110607659-004

much shorter periods of qualitative changes which frequently take the form of truly Cambrian explosions, the practical results of which only impose themselves little by little. As an academic and scientific discipline as well as a millenarian cultural practice, lexicography is now experiencing such a moment with many new ideas and visions that still have to be tested and put into practice, cf. Fuertes-Olivera (2016).

2 Current situation and challenges

As any other discipline that contains a strong element of artisanship in its practical dimension, lexicography is highly influenced by the development of technology. It is easy to observe this close and also complex relationship between lexicography and the technology that is available at a given moment. In a historical perspective, the current situation within the discipline has been triggered by the vertiginous development of the technological basis of our society during the past few decades and the introduction of new computer, information and communication technologies and techniques which have materialized in computers, databases, Internet, information clouds, and many other inventions. The new technologies have penetrated and thoroughly shaken practical lexicography in its three basic components, namely the compilation, presentation and usage of dictionaries and other lexicographical products. In a certain way, the millenarian discipline is shaken by an identity crisis due to this transition from the well-known to the unknown, cf. Tarp (2016a).

The very organisation of the lexicographical work is changing completely with still more dictionaries compiled in the clouds from work stations connected to remote databases. New methods of selecting, storing and presenting the lexicographical data have been introduced together with new ways of satisfying the users' needs. Many dictionary projects rely on gigantic text corpora or obtain their data directly from the immense material made available on the Internet. Part of the compilation process has been automatized by means of advanced computer programs at the same time as alternative, collaborative compilation methods are appearing, cf. Rundell/Kilgarriff (2011) and Rundell (2015).

The printed dictionary is gradually being relegated to the museum of antiquities. The tendency is global and seems irreversible. Most digital dictionaries are now published online, i.e. in the clouds, and can be accessed almost anywhere on the earth by means of a variety of devices like laptops, tablets and smart phones. Simultaneously, the traditional stand-alone dictionary aimed at the user's decisive consultation is being replaced, at least partially, by lexicographical products that are integrated into other types of digital tools designed to assist the writing, translation and reading of texts or the learning of languages and other areas of human knowledge and activity, cf. Tarp (2018b).

Although we are still in the middle of the technological transition period, there are already more than enough elements that allow us to conclude that the current changes are much deeper and more far-reaching than those which took place after the introduction of the printing technology, in Europe about 500 years ago and in China and Korea various centuries before, cf. Hanks (2010, 2013).

It is, however, worth noting that the new technologies do not only offer new methods and better solutions but also create new problems and challenges, frequently due to a too enthusiastic and ill-considered use of them. For instance, the big and understandable enthusiasm created by the *big data* –i.e. the gigantic amount of data accessible on the Internet or in big corpora stored in databases– sometimes makes the lexicographers forget the *smart data* adapted to the specific needs of their users. The undesirable result of this exaggerated focus on quantity is information overload which chases away many users who instead try to solve their problems by means other information sources, cf. Tarp/Gouws (2017).

Another and even more serious challenge are the many internauts, especially young people, who have become used to free data access and for this reason, or simply because they have no money, prefer to consult the growing number of free online dictionaries which are more often than not of a dubious quality. As a result, and due to the lack of an appropriate business model, many publishing houses are struggling with still bigger financial problems due to reduced sales and several of them, even some well-known and prestigious ones, have already thrown the towel into the ring and closed down their sections for production and commercialization of dictionaries. Consequently, the balance is tipping still more in favour of dictionaries of dubious quality in the detriment of high-quality dictionaries.

The problem is real and must be taken seriously. It does not only affect individuals but society as a whole and is co-responsible for a growing number of badly written and translated texts as well as many misunderstandings which complicate human life unnecessarily and represent an extravagant waste of time and money. In fact, the current situation is full of paradoxes:

- On the one hand, there are more users of dictionaries than ever before, a fact reflecting success. On the other hand, there are more potential users of dictionaries not consulting them than ever before, a fact expressing crisis.
- On the one hand, many publishers of high-quality dictionaries have been forced to close down their business due to dramatically reduced sales. On the other hand, low-quality online dictionaries with free access flourish like mushrooms after the rain.
- On the one hand, modern information-age users need high-quality dictionaries providing quick and reliable information to solve their complex problems and needs. On the other hand, a growing number of these users opt for free-access dictionaries of dubious quality frequently obtaining inadequate and even incorrect information which only adds to their problems.

The only positive aspect in this complex of paradoxes is that users get access to a big amount of lexicographical material, part of which would not have been published otherwise although it may be of big value to at least some of them. However, the negative aspect is that reliable and unreliable dictionaries are mixed, leaving many users in a position where they are not able to determine the quality of the lexicographical data and the retrieved information. In both cases, dictionary users are increasingly losing their innocence as still more –and occasionally too much– is required from them in terms of their critical sense, lexicographical skills and ability to access relevant data and filter the retrieved information.

There seems to be no short-term solution to all these problems and challenges as they are closely related to the current social and economic structures of our societies. Apart from public funding of projects most relevant to society and the development of a sustainable business model in the publishing houses, one step that could bring some relief would undoubtedly be to make better use of the new technologies and techniques in order to develop more productive, and hence more economical, compilation methods without compromising the quality of the final product. In order to set out on this road, a critical and self-critical vision of current lexicographical work is required together with a considerable dose of the "audacity, audacity, audacity" which Danton called for during the great French Revolution.

The problem is above all subjective as technology in itself by no means can be blamed for the calamities. The new technologies and techniques have mainly been developed as a creative response to needs observed within other areas of activity and have until now only partially been transferred to lexicography, seemingly in a chaotic and deformed way. Here the problem is the general lack of visions and ideas among many lexicographers and the stubborn tendency to do business as usual instead of taking full advantage of the new possibilities. In this respect, it is worth remembering the words which the Cuban philosopher and poet José Martí addressed to his country's rebellious youth in 1891: "There is too much imitation, and creation holds the keys to salvation. Creation is the password of this generation".

3 Going to the roots

A major problem in current lexicography is its performers' self-understanding, i.e. what they understand by lexicography and how they view it in its interaction with other disciplines. Regretfully, many lexicographers still consider their discipline to be a sub-discipline of linguistics, the purpose of which is to describe language and present this description in the form of dictionaries. By doing so, they ignore a very big number of dictionaries, maybe the majority, which do not deal with language or linguistics but with other areas of human activity and knowledge, e.g. dictionaries of economics, geography, history and many other disciplines and sciences. These lexicographers, many of whom have their academic background in linguistics, do not go to the core of their discipline which is to provide *assistance* to users with *information needs* that may occur in different situations *within any area of human activity and knowledge*. In this respect, lexicography can be defined as an information discipline in its own right, but with a big interdisciplinary collaboration, cf. Tarp (2018a).

With such an understanding of lexicography, it is much easier to determine what is common to all dictionaries and which principles should govern the interaction with all other disciplines related to the production of dictionaries. This approach is especially relevant when the hierarchy between lexicography and technology has to be established. Figure 1 shows the different types of knowledge and competences needed in a modern online dictionary project based on the idea that lexicography is an independent science characterised by a big interdisciplinary vocation.

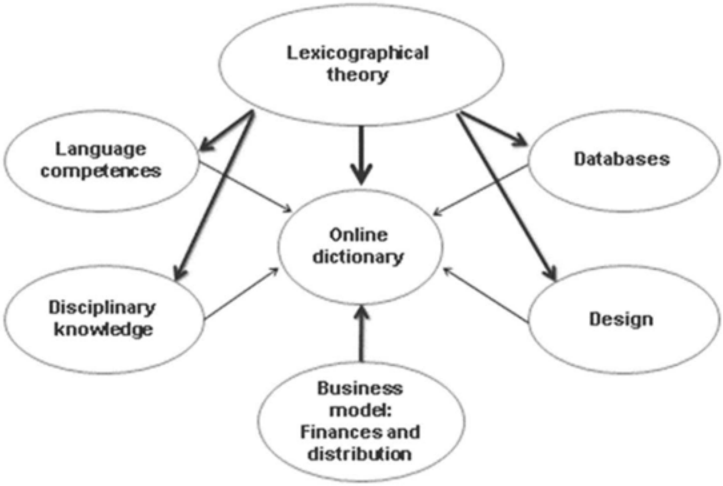


Fig. 1: Knowledge and competences required in an online dictionary project

Figure 1 suggests that lexicographical principles should guide all the other types of knowledge and skills needed in a specific online dictionary project. By language competence is here understood the competence which any educated adult person is expected to have in his or her own language (and another language when necessary) and which is required when writing definitions and preparing other types of lexicographical data. By disciplinary knowledge is understood the expert knowledge of the topic(s) treated in the dictionary, e.g. commerce, economics, geography, biology or linguistics, whether it is a general or a specialised one, cf. Bergenholtz (2013).

The lexicographers doing the practical work should adapt their knowledge and competences to the lexicographical principles guiding the project in order to guarantee the quality of the dictionary and prepare definitions and other types of data that are useful and easily understandable to the foreseen user group. The same can be said about the knowledge and skills required to prepare lexicographical databases and other tools to be used in the compilation process as well as to design appropriate user interfaces both for the lexicographers (data input) and the end users (data output). The lexicographer in charge of a project is not supposed to perform this part of the job, but he or she should have a minimum of knowledge about databases, interfaces and other relevant tools that allows for a fruitful communication with the programmer or designer and the formulation of lexicographical instructions to these essential collaborators for the benefit of the final product, cf. Tarp (2015).

All this implies that the *overall management of dictionary projects* should be the responsibility of someone who is trained in lexicographical theory and practice and possesses a minimum of relevant technological knowledge, cf. Bergenholtz (2018).

The only exception to this rule is the financial side of the project. No big dictionary project can be carried out and finished without money, and frequently it is easier to climb Mount Everest than finding an interested sponsor or investor. The necessary funding may come from the state, a university, a private foundation, an enterprise or a publishing house (that has developed a sustainable business model), or a combination of these. However, even when available, the money is frequently scarce and sets limits to the project in terms of the technology and techniques to be applied and the amount of lexicographical data to be included. Sometimes the sponsor or investor may also have special demands which may have negative impact on the project from a lexicographical point of view. All this requires pragmatism and the ability to make compromises but it does not change the basic idea, namely that project management in the narrow sense of the word should be in the hand of a trained lexicographer, and not in the hand of an investor, a programmer or anybody else who only knows his or her own discipline well, e.g. linguistics, but not lexicographical theory and practice.

4 Theory and methodology

Just as there are disagreements in the community about the disciplinary status of lexicography and its scientific character, there are also disagreements concerning the existence and possibility of developing lexicographical theories. This discussion has reached the absurd point where there still are some researchers, among them

Bejoint (2010), who close their eyes to reality and try to maintain the discipline at the level of a traditional "art-and-craft" philosophy, cf. Tarp (2016b).

It goes without saying that such an approach to lexicography is not helpful in the current situation where the crucial decisions and steps to be taken more than anything else require guidance from an advanced theory that is based on careful observation of practice. Function Theory is such a theory although it does not claim to represent the only road leading to Rome. This theory states that the very essence of dictionaries is that they are conceived as a response to information needs observed in society, and that the specific content of these needs is determined by the social context or situation where they occur as well as the relevant characteristics of the person who experiences them. In this perspective, lexicography is understood not only as an information discipline but also as a social discipline. The vision implies that dictionaries may have different functions according to the types of information needs they are designed to meet, cf. Tarp (2008) and Fuertes-Olivera/Tarp (2014).

Function Theory does not by itself solve the current challenges to lexicography but it provides reasonable guidance to the development of a methodology that can take the discipline forward. This methodology encompasses four general principles as well as a series of specific methods that can be applied in order to determine relevant user needs and lexicographical data and to select, prepare and present these data, cf. Tarp (2014).

According to the proposed methodology, the first step in a dictionary project is always to establish the function(s) of the planned dictionary or set of dictionaries, i.e. the social situation(s) to be covered by the project as well as the intended user group and its possible needs in these situations. When this is done, the next step is to determine the relevant user needs as well as the lexicographical data categories which, apart from the compulsory lemmata, could be part of speech, definition, morphological and syntactic data, synonyms, antonyms, example sentences, illustrations, links, etc. Different from projects based on other principles, this is normally done using the method of deduction, a very fast method that requires both lexicographical and topical (subject-field) knowledge and which has been described in details by Fuertes-Olivera/Tarp (2014).

When the data categories to be included in the project have been established, the lexicographer in charge of the project has to decide how these categories should be presented to the end users, for instance, if they should be grouped according to functions or with the possibility of further individualising the access process. Based on this decision, an analysis of all the data categories is made with a view to establishing the smallest elements that may be displayed individually. This is required as an essential part of the preparation of another important step, namely the design of the lexicographical database.

5 Lexicographical databases

Every lexicographical project has its own personality. Although there are various dictionary writing systems available on the market, the best option is always to prepare a new database from scratch for each new project. This is, at least, the experience obtained from a number of general and specialised dictionary projects carried out at the Centre for Lexicography in Aarhus (Denmark) and the International Centre of Lexicography in Valladolid (Spain), both of which work on the basis of Function Theory. The design of a new database adapted to the project in question is important, not only to guarantee the quality of the final product, but also to ensure that the compilation process can be completed as smoothly and fast as possible.

The design of a lexicographical database is the task of a skilled programmer but it is the responsibility of the lexicographer in charge of the dictionary project to provide the pertinent instructions. In this respect, what basically has to be communicated to the programmer is the number of fields to be included in each data card as well as their mutual relations. The fields may include 1) fields for each category of data to be displayed individually, 2) reserve fields to be used if the need for more data categories is discovered later in the process, and 3) special memo fields to be used by the lexicographers, cf. Bergenholtz/Nielsen (2013).

The number of fields varies from project to project. Bergenholtz/Agerbo (2015), for instance, report 24 specific data fields included in the database feeding *De Danske Netordbøger* (Danish Internet Dictionaries) which consist of a series of seven monolingual dictionaries. Based on the same methodology, Fuertes-Olivera (2015) lists 18 fields in the database sustaining the Spanish-English-Spanish *Diccionarios de Contabilidad* (Accounting Dictionaries), to which should also be added reserve and memo fields. And according to Bergenholtz/Fuertes-Olivera (2015), the following 30 fields can be found in the database designed to be used in the compilation of the *Diccionarios en Línea de Español* (Spanish Online Dictionaries) which also consist of various monolingual dictionaries:

Ex. 1: The entry house in The Logos Dictionary (Excerpts)

Lemma
Style note to lemma
Sublemma
Style note to sublemma
Homonym number
Polysemy number
Meaning
Lexical remark
Usage remark
Part of speech

Grammar, recommended inflexion
Grammar, non-recommended inflexion (one or more)
Grammar, spelling remark
First reference
Second reference(s)
Collocation(s)
Example(s)
Word formation(s)
Synonym(s)
Style note to synonyms
Antonym(s)
Style note to antonyms
Synonym remark
Proverb(s)
Idiom(s)
Idiom meaning
Internet link
Grammatical reference
Grammar (valency) of idioms
Field for communication among the lexicographers

These 30 data fields were also those (together with reserve fields) which were originally communicated to the programmer who designed the database to be used in the *Diccionarios en Línea de Español*. The plural forms used to name some of the fields indicate that the corresponding fields can be repeated, e.g. when there are more than one collocation, idiom, etc. As can be seen, no specific relations between the fields are yet indicated in the above list as this generally requires small work sections with oral communication between programmer and lexicographer, cf. Tarp (2015).

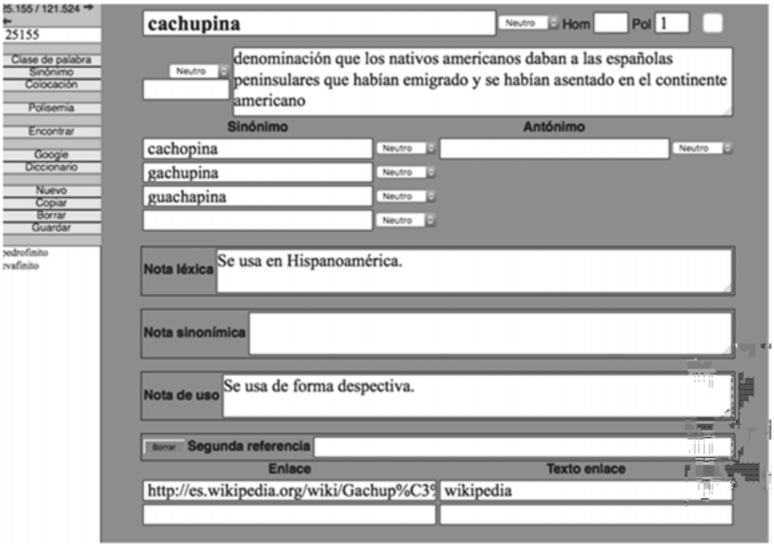
It is important to stress that the order of the 30 data fields listed above is arbitrary and does in no way represent the structure of the online dictionary articles presented to the users. In fact, by means of modern information techniques it is possible to construct various, completely different dictionaries based on one and the same database as is the case with the Spanish Online Dictionaries. We will return to this question in Paragraph 8.

6 Means of production

A dictionary is a product that frequently takes the form of a commodity. As such, its compilation is a production process just as any other process where goods are produced. This specific process should be performed by knowledgeable and skilled lexicographers using means of production, i.e. tools or software with which they process the raw material (raw data) and make it available and useful to the consumers of this product. Hence, apart from the database, the lexicographical means of production entail a number of special tools with which the lexicographers search and select the relevant raw data as well as the interface which they use to introduce the processed data into the database.

The most important means of production is the *lexicographer's interface*. As was the case with the database, this interface should also be designed especially to each dictionary project and taking into account the specific tasks to be performed and the concrete persons involved. In this respect, various important requirements have to be fulfilled in order to guarantee both high quality and high productivity. Technically, this requires that it contains all the fields needed to introduce lexicographical data into the database. But the interface should also be as easy and comfortable as possible to work with in order to facilitate the lexicographer's job, reduce the number of mistakes, economise on the resources employed, and shorten the total production time. In addition to the mentioned data fields, the well-designed interface should therefore also contain a number of functional buttons which allow the lexicographers to perform searches for raw data, work with the database, navigate on the different pages of the interface and introduce standardised lexicographical data like part of speech, inflection forms and style notes in the database, cf. Bergenholtz/Nielsen (2013: 83).

In most dictionary projects the number of fields and buttons is too large to fit into a single screen page. In order to avoid scrolling down or switching to the sides, actions which disturb the lexicographer's overview, it may thus be necessary to divide the interface into various pages according to criteria that are determined by the characteristics of each specific compilation process. In dictionary projects where several lexicographers work together with different and clearly defined tasks, a separate page could, for instance, be prepared to each of them in order to facilitate the work and avoid, or at least reduce, navigation between pages. Apart from distributing the data fields and buttons on different pages, it is also necessary to find a proper design of the individual pages.



25155

cachupina

Pol 1

denominación que los nativos americanos daban a las españolas peninsulares que habían emigrado y se habían asentado en el continente americano

Sinónimo	Antónimo
cachopina	
gachupina	
guachapina	

pedrofinito
evafinito

Nota léxica: Se usa en Hispanoamérica.

Nota sinonímica:

Nota de uso: Se usa de forma despectiva.

Segunda referencia:

Enlace	Texto enlace
http://es.wikipedia.org/wiki/Gachup%C3%	wikipedia

Ex. 2: Lexicographer's interface designed to the Spanish Online Dictionaries

Example 2 is a screen shot showing the interface used by the lexicographers to introduce data into the database sustaining the Spanish Online Dictionaries mentioned above. The screen shot shows the page where definitions, synonyms and antonyms are introduced to one of senses of the polysemous Spanish noun *cachupina*. As can be seen, the interface has been designed with blue and green colours which are pleasant to most people's eyes. In the upper part of the page there are fields to indicate style, homonymy and polysemy (the polysemy number indicates that the page belongs to the card representing the first sense of the word). In the centre of the page, one can find the main data fields where the definition, synonyms, antonyms as well as lexical and usage remarks can be introduced. To the left there is a column with a number of buttons where the lexicographer can navigate between the pages belonging to the card, work with the database, and perform searches on the Internet by means of Google. Finally, below the column there is a field for internal communication among the lexicographers.

Too many interfaces in the current dictionary writing systems still have a very primitive and, up to a certain point, chaotic design which makes them more difficult to work with and raises the risk of mistakes. The result may be lower quality and, especially, much longer production time. By contrast, the lexicographers working on the Spanish Online Dictionaries with the above interface, as well as the methods developed to collect and process data, are able to complete 4-6 senses (cards) per hour, including the selection of meaning elements, writing of definitions and preparation of all other relevant lexicographical data, cf. Tarp/Fuertes-Olivera (2016).

7 Lemma selection and frequency

In a very informative paper where Rundell/Kilgarriff (2011) discuss the relation between dictionary compilation and computer technology during the past five decades, they formulate the following basic principle that should govern the selection of lemmata for a new dictionary project: “Building a headword list is the most obvious way to use a corpus for making a dictionary. *Ceteris paribus*, if a dictionary is to have N words in it, they should be the N words from the top of the corpus frequency list.” (Rundell/Kilgarriff 2011: 263).

This idea is shared by many other lexicographers. However, as everything has to be questioned in the current transition period, the question has to be asked whether corpus frequency is indeed the best guiding principle for lemma selection to a dictionary covering general language? It may have been, even when the two authors published their article, but is it still the case?

In 2012, Bergenholtz and Norddahl published, regretfully only in Danish, a research article with very surprising findings. The two scholars, a lexicographer and a programmer, had discovered that only one third of the more than 110,000 lemmata contained in the Danish Internet Dictionaries had been consulted several years after their publication. This implied that two thirds of all lemmata had never been looked up although the total number of consultations exceeded 19 million, a tendency that has been stable since then. The two authors also found out that various words with high frequency on the Internet had not been consulted even one single time, whereas other low-frequency words had been looked up several times, cf. Bergenholtz/Norddahl (2012).

These findings indicate that there is no direct relation between a word’s frequency in a corpus (in this case the Internet) and the dictionary users’ interest in or need for consulting it. This discovery brings us back to the discussion of lexicography’s disciplinary status and its relationship to linguistics. If the purpose of a dictionary is to describe language, it makes perfect sense to include the most frequent words into it. By contrast, if the purpose is to assist users with information needs, it would be more appropriate to include the words which the users, for one reason or another, find troublesome and therefore tend to consult. In the last case, word frequency is no longer related to a corpus but to problems and needs. It should therefore be the most troublesome and frequently consulted words that top the list of lemma candidates.

Hence, the challenge is to identify the words which create (most) problems for the intended users of a dictionary. It is not as easy as it sounds. Bergenholtz/Norddahl (2012) analyzed the non-consulted words from various angles and, with the exception of a few old or very new words, no systematic explanation to this strange phenomenon could be found. The words in question belonged to all parts of speech; they were single and compound words, old and new, native and imported ones. The two researchers found no indications whatsoever that explained why the users had consulted some words, occasionally very often, while other words were completely ignored although they had a high-frequency score on the Internet.

So, what is to be done? In some cases nothing can probably be done, at least for the time being. Besides, for big general-language dictionary projects the problem mentioned may not be so relevant as the challenge here is to collect as many lemmata as possible. This is due to the almost unlimited storage capacity of the database which makes it relevant to change the old principle where the lexicographers should justify the inclusion of a new word, with a new principle which requires that they justifies its exclusion, cf. Rundell (2015: 312).

However, as the 2016 Nobel Prize winner Bob Dylan sang in his youth: "The Times They Are A-Changin'". During the past few years an increasing number of dictionaries of very different types have been placed on the Internet. These dictionaries have been consulted by millions of users who have left their tracks and behaviour recorded in log files. Based on these log files it is quite easy to establish which words have most frequently –or never– been consulted in each type of dictionary. In this perspective, log files make up an important empirical source for dictionary compilation which still has to be explored fully. The technology is already there, it just needs to be used appropriately and visionarily.

8 Needs-adapted data presentation

A modern online dictionary is a never ending saga. It can be published in its first "edition" when a certain amount of lemmata and addressed lexicographical data are ready. From then on the publication can be made continuously, and the same holds true for the subsequent updatings which, at least in principle, can go on forever as a fluid response to abortive lookups and other types of user behaviour as well as the normal development of the topic(s) or language(s) treated in the dictionary. As we saw above with the log files, today the possibilities of interaction with the users are bigger than ever before. It is therefore sad to see the low quality of many current online dictionaries.

In a certain sense we can speak of *lexicographical alienation*. In its origins, the discipline was born and developed as a response to needs detected in society. Lexicographers were very close to their users whom they often knew personally. As it also happened within other social disciplines, little by little lexicography grew more introverted. The close relationship with the users and the knowledge of their needs were diluted by lexicographical habits, techniques and principles, frequently taken in from other disciplines. The user became still more distant and ended up being the "well-known unknown" (Wiegand 1977: 59).

The problem may not have been so dramatic in the relatively long period when dictionaries developed within the old printed paradigm and where the possibilities of providing targeted user service were limited by the book format and the market conditions. This situation changed completely with the advent of the new disruptive technologies. Now it suddenly became possible, not only to give a targeted but also a still more personalised or individualised assistance to the users of dictionaries, cf. Rundell (2010) and Tarp (2011).

Bothma (2011) presents various information techniques which have been developed in the framework of information science and which, according to him, may have relevance for lexicography. Some of these techniques have already been applied in a number of dictionary projects; others are still to be tested while a few may not be relevant, at least for the time being. The most interesting techniques seem to be *data filtering* –which allows the lexicographers to present data to users with specific types of needs, even on delivery– as well as *adaptive representation* and *reuse of data through linking*, thus coming closer to a more flexible, dynamic and personalised lexicographical product which, as an extra gift, is happily free from information overload.

However, it is somehow thought-provoking that of the information techniques discussed by Bothma, it is generally the two latter that have been applied in recent dictionary projects, i.e. the techniques that can be used to adapt the dictionary article to the screen and access supplementary data by means of indexing, hypertexts, hidden data, pop-up windows, database-internal links, links to the Internet or external corpora, etc. Surprisingly, it seems that only few dictionaries use data filtering through user profiling and description of the situation where they experience their needs, e.g. text reception, production and translation.

Examples of dictionary projects that have, at least partially, used this technique are the *Danske Netordbøger*, the *Accounting Dictionaries* and the *Diccionarios de Contabilidad*, all of which consist of a series of dictionaries adapted to the needs which the users may have in specific contexts. Each of the three projects mentioned is sustained by a single database and encompasses a set of different user interfaces, each of which represents a separate dictionary and exclusively furnishes lexicographical data to support the function of this dictionary, thus filtering the data stored in the database. These dictionaries, however, do not use some of the other techniques mentioned by Bothma (2011), e.g. some of the adaptive techniques shown by Rundell (2015).

In a frequently quoted contribution, de Schryver (2003) wrote about the lexicographers' *dreams* in the age of digital dictionaries. However, much more than dreams which always include the risk of falling asleep, what is needed today more than anything else is *visions*, i.e. the ability to connect lexicography and technology in an ever ascending spiral that constantly provides new solutions to both old and new problems.

A fundamental element in the new visions that ought to dominate lexicography in the nearby future is that the old *static approach* where the dictionary articles and lexicographical data are identical for all users in all consultations should be replaced by a *dynamic approach*. According to this approach, the articles and data displayed on the screen should vary from consultation to consultation and adapt to the specific needs, not only of a determined user group but increasingly also of individual users. This provision of a still more personalised service corresponds to the general trend in our societies:

> The demand among consumers generally is for products that match their individual needs more precisely – an expectation that is already transforming businesses like television and popular music, for example. In dictionary terms, this implies both customization and personalization.
>
> Rundell 2010: 172

9 Conclusion

Although the somehow provocative title of this contribution suggests that the combination of new technology and lack of vision is a dangerous cocktail, we have nevertheless tried to transmit some of the positive aspects of modern lexicography and show that there are solutions to all challenges, and many more than the ones discussed in the previous paragraphs. However, this does not imply that we should forget the many negative aspects, of which some were listed in the beginning of the contribution.

There is still a reluctance to accept lexicography as a discipline in its own right and to work with evidence-based theories as, for instance, Function Theory. There is a tendency to only half-heartedly embrace the new technologies and techniques and to approach online dictionaries as were they still paper-based. There are still too many lexicographical databases, and especially interfaces for users and lexicographers, that seem to have been designed in the dinosaur-age of modern computer technology. There is still too much business-as-usual and too few visions. In this respect, it is worth once more repeating the instructive words with which Gouws (2011) told his colleagues that it is not enough to learn, we also have to “unlearn”, i.e. get rid of old habits and ways of thinking:

> Looking back at the development of the theory and practice of lexicography it is clear that for too long the practice of printed dictionaries had to go without a sound theory, for too long lexicography did not establish itself as an independent discipline, for too long the pool of lexicographers had been restricted to experts from a single field, for too long innovation in the lexicographic practice was impeded by its theory being a follower and not a leader, for too long lexicographic theory was exclusively directed at being implemented in the production of dic-

> tionaries. Looking at the future, the planning and compilation of electronic dictionaries and the further development of a coherent and medium-unspecific theory we need to unlearn a lot, we need to learn a lot so that we can be innovative and produce better reference tools, including even dictionaries.
>
> Gouws 2011: 29

10 Bibliography

10.1 Dictionaries

Nielsen, Sandro/Mourier, Lise/Bergenholtz, Henning (2016): Accounting Dictionaries. Odense: [<Ordbogen.com>].

Bergenholtz, Henning, in cooperation with Heidi Agerbo, Heidi/ Michelsen, Andreas Bock/Eriksen, Kathrine Brosbøl/Bodilsen, Andreas/Gudmann, Helene R./Koed, Aleksander/Poulsen, Jon Nørgaard/ Nielsen, Mia Lybkær Kronborg/Nguyen, Jane/Thers, Henrik (2015): De Danske Netordbøger. Database: Richard Almind and Martin Carlsen. Odense: [<Ordbogen.com>].

Fuertes-Olivera, Pedro A./Bergenholtz, Henning/Nielsen, Sandro/Gordo Gómez, Pablo/Niño Amo, Marta/Ríos Rodicio, Ángel/Sastre Ruano, Ángeles/Tarp, Sven/Velasco Sacristán, Marisol (2013): Diccionario Español de Contabilidad. Base de Datos y Diseño: Richard Almind y Jesper Skovgård Nielsen. Hamburg: [<Lemma.com>].

Fuertes-Olivera, Pedro A./Bergenholtz, Henning, in collaboration with Sastre Ruano, María Ángeles/Álvarez Ramos, Eva/Fonseca Herández, María/López Carrero, María José/Prieto Salvador, Álvaro/Saldaña, Olga (under construction): Diccionarios en Línea de Español Valladolid-UVa. Hamburg: [<Lemma.com>].

10.2 Monographes and articles

Bejoint, Henri (2010): The lexicography of English. Oxford: Oxford University Press.

Bergenholtz, Henning (2013): The Role of Linguists in Planning and Making Dictionaries in the Modern Information Society. In: Kwary, Deny A./Wulan, Nun/Musyahd, Lilla (eds.): Lexicography and Dictionaries in the Information Age. Surabaya: Airlangga University Press, 1–19.

Bergenholtz, Henning (2018): Dictionary Management. In: Fuertes-Olivera, Pedro A. (ed.): Routledge Handbook of Lexicography. London/New York: Routledge, 34–42.

Bergenholtz, Henning/Agerbo, Heidi (2015): Lexicographical structuring: the number and types of fields, data distribution, searching and data presentation. In: Lexicographica 31, 5–37.

Bergenholtz, Henning/Fuertes-Olivera, Pedro A. (2015): Los Diccionarios en Línea de Español "Universidad de Valladolid". In: Estudios de Lexicografía 4, 71–98.

Bergenholtz, Henning /Norddahl, Bjarni (2012): Ordbogsartikler som ingen læser. In: LexicoNordic 19, 207–223.

Bergenholtz, Henning/Nielsen, Jesper Skovgård (2013): What is a Lexicographical Database? In: Lexikos 23, 77–87.

Bothma, Theo J.D. (2011): Filtering and Adapting Data and Information in the Online Environment in Response to User Needs. In: Fuertes-Olivera, Pedro A./Bergenholtz Henning (eds.): e-Lexicography: The Internet, Digital Initiatives and Lexicography. London/New York: Continuum. 2011, 71–102.

De Schryver, Gilles-Maurice (2003): Lexicographers' dreams in the electronic-dictionary age. In: International Journal of Lexicography 16 (2), 143–199.

Fuertes-Olivera, Pedro A. (2015): La Lexicografía de Internet: Los Diccionarios de Contabilidad. In: Estudios de Lexicografía 1, 49–59.

Fuertes-Olivera, Pedro A. (2016): A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. In: International Journal of Lexicography 29 (2), 226–247.

Fuertes-Olivera, Pedro A./Tarp, Sven (2014): Theory and Practice of Specialised Online Dictionaries: Lexicography versus terminography. Berlin/Boston: De Gruyter.

Gouws, Rufus H. (2011): Learning, Unlearning and Innovation in the Planning of Electronic Dictionaries. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography: The Internet, Digital Initiatives and Lexicography. London/New York: Continuum, 17–29.

Hanks, Patrick (2010): Lexicography, Printing Technology, and the Spread of Renaissance Culture. In: Dykstra, Anne/Schoonheim, Tanneke (eds.): Proceedings of the XIV Euralex International Congress. Leeuwarden: Fryske Akademy, 988–1016.

Hanks, Patrick (2013): Lexicography from Earliest Times to the Present. In: Keith Allan (ed.): The Oxford Handbook of the History of Linguistics. Oxford: Oxford University Press, 503–536.

Martí, José (1891): Nuestra América. In: La Revista Ilustrada de Nueva York. 10th January.

Rundell, Michael (2010): What future for the learner's dictionary? In: Kernerman, Ilan J./Bogaards, Paul (eds.): English Learners' Dictionaries at the DSNA 2009. Jerusalem: Kdictionaries, 169–175.

Rundell, Michael (2015): From Print to Digital: Implications for Dictionary Policy and Lexicographic Conventions. In: Lexikos 25, 301–322.

Rundell, Michael/Kilgarriff, Adam (2011): Automating the creation of dictionaries: where will it all end? In: Meunier, Fanny/De Cock, Sylvie/Gilquin, Gaëtanelle/Paquot, Magali (eds.): A Taste for Corpora. In honour of Sylviane Granger. Amsterdam/Philadelphia: Benjamins, 257–282.

Tarp, Sven (2008): Lexicography in the borderland between knowledge and non-knowledge. Tübingen: Max Niemeyer.

Tarp, Sven (2011): Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs Satisfaction. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography: The Internet, Digital Initiatives and Lexicography. London/New York: Continuum, 54–70.

Tarp, Sven (2014): Theory-Based Lexicographical Methods in a Functional Perspective: An Overview. In: Lexicographica 30, 58–76.

Tarp, Sven (2015): Structures in the Communication between Lexicographer and Programmer: Database and Interface. In: Lexicographica 31, 17–46.

Tarp, Sven (2016a): Challenges to Lexicography in the Digital Era: the Point of View of the Function Theory. In: Develi, Hayati/Gürlek, Mehmet (eds.): I. ve II. Uluslararası Sözlükbilim Sempozyumu Bildiri Kitabı. Istanbul: İstanbul Metropolitan Municipality Kültür A.Ş, 13–35.

Tarp, Sven (2016b): The Amazing Vitality of Things that Don't Exist. In: Schierholz, Stefan/Gouws, Rufus H./Hollós, Zita/Wolski, Werner (eds.): Wörterbuchforschung und Lexikographie. Berlin/Boston: De Gruyter, 227–237.

Tarp, Sven (2018a): Lexicography as an Independent Science. In: Fuertes-Olivera, Pedro A. (ed.): Routledge Handbook of Lexicography. London/New York: Routledge, 19–33.

Tarp, Sven (2018b): The Concept of Dictionary. In Pedro A. Fuertes-Olivera (ed.): Routledge Routledge Handbook of Lexicography. London/New York: Routledge, 237–249.

Tarp, Sven/Fuertes-Olivera, Pedro A. (2016): Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-Uva. In: Lexikos 26, 273–296.

Tarp, Sven/Gouws, Rufus H. (2017): Information Overload and Data Overload in lexicography. In: International Journal of Lexicography 30 (4), 389–415.

Wiegand, Herbert Ernst (1977): Nachdenken über Wörterbücher. In: Drosdowski, Günther/Henne, Helmut/Wiegand, Herbert E. (eds.): Nachdenken über Wörterbücher. Mannheim/Wien/Zürich: Bibliographisches Institut, 51–102.

Section 2: **Multilingual electronic dictionaries**

Olga Batiukova and Elena de Miguel

Multilingual Electronic Dictionary of Motion Verbs (DICEMTO): overall structure and the case of andar

Abstract: This chapter presents the research project "Multilingual Electronic Dictionary of Motion Verbs (DICEMTO)", focused on the meaning alternations displayed by predicates with motion verbs. The basic assumption underlying DICEMTO is that the semantic variation that these verbs show in context is determined by the sub-lexical features encoded in their lexical entry and the lexical entries of their arguments. We explore in detail the lexical entry of the Spanish verb andar, composed of the minimal definition, the argument structure, the event structure, and the qualia structure, and show how the different meaning components belonging to these levels interact dynamically and generate different senses depending on the context.

Keywords: Generative Lexicon, qualia, motion verbs, event structure

1 Introduction

This paper presents the research project "Multilingual Electronic Dictionary of Motion Verbs (DICEMTO)"[1], which has been developed by the research group UPSTAIRS (Unit for the Word Study: Internal Structure and Syntactic Relationships) at the Department of Spanish Language and Literature, Autonomous University of Madrid[2].

This project is focused on motion verbs, and its main goal is to register in a systematic and consistent way the different senses that these verbs display in context

[1] The research project presented in this paper has been financed by the Ministry of Economy and Competitiveness of Spain under research grants FFI2009-12191 and FFI-2012-33807 (Principal Investigator: Elena de Miguel). The website of the research group UPSTAIRS can be accessed at [<https://www.uam.es/gruposinv/upstairs/diccionario.html>].

[2] A full list of the researchers involved in this project, including collaborators from other Spanish and foreign universities, can be found at [<https://www.uam.es/gruposinv/upstairs/upstairs2/agentivo.html>].

Olga Batiukova: Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid,
tel. +34 914972023, fax +34 914974184, volha.batsiukova@uam.es
Elena de Miguel: Universidad Autónoma de Madrid, Campus de Cantoblanco, 28049 Madrid,
tel. +34 914974504, fax +34 914974184, elena.demiguel@uam.es

https://doi.org/10.1515/9783110607659-005

as a result of their combination with other components of the predicate. Motion verbs were chosen of all the semantic groups of verbs due to their capacity of lightening and extending their basic spatial meaning when forming verbal periphrases and idiomatic expressions in different natural languages, independently of their typological ascription. DICEMTO is a theoretical dictionary rather than a conventional one: in looking into the combinatorial potential of motion verbs, we align ourselves with the frameworks that seek to identify the lexical features that determine the syntactic behavior of words and that ultimately license certain meaning alterations in the form of derived verb senses, idiomatic expressions, and verbal periphrases.

2 Theoretical foundations and general design of the dictionary

DICEMTO was conceived as a theoretical linguistic project. It is our belief that, in order to effectively fulfill their main function (i.e., explaining word meaning), the dictionaries must acknowledge the advances in lexical semantics and lexicon-syntax interface that allow accounting for the syntactic and semantic behavior of lexical units (cf. Batiukova 2009b). Lexicographical and computational projects WordNet, VerbNet, FrameNet, Redes and ADESSE[3] are some of the most representative initiatives following this approach.

We assume the existence of a set of grammatically-relevant lexical features and generative mechanisms that determine in what contexts a word can appear and what meaning extensions (both literal and apparently figurative) it can display depending on the context. In terms of Generative Lexicon theory (henceforth GL), some of these lexical-semantic features make up specific sublexical structures. As will be shown in sections 3.2.1–3.2.3, the Event Structure and the Qualia Structure, among others, provide bits of information that ultimately license different word combinations, both apparently free and apparently constrained. We also make use of concepts routinely adopted in other influential linguistic models (generative grammar, frame semantics, etc.), such as argument structure, thematic structure and semantic roles (see section 3 for details).

[3] WordNet = [<https://wordnet.princeton.edu/>].
VerbNet = [<https://verbs.colorado.edu/~mpalmer/projects/verbnet.html>].
FrameNet = [<https://framenet.icsi.berkeley.edu/fndrupal/>].
ADESSE = [<http://adesse.uvigo.es/data/>].
All the links were accessed online on July 15, 2016.

One of the main achievements of this project is the design of a linguistic meta-entry, which serves as a template for the lexical entries of concrete motion verbs and which can be easily adapted to other groups of verbs. This meta-entry is an abstract model based on a detailed study of particular motion verbs, and it serves, in turn, to systematize and improve the lexicographic representation of concrete verbs.

At this point DICEMTO has ten full lexical entries for the verbs andar, bajar, caer, entrar, ir, llegar, salir, subir, venir, and volver. The main language of the dictionary is European Spanish: all the lexical entries are written in Spanish and the primary examples are provided in Spanish. These examples are further translated into fourteen other languages: Arabic, Armenian, Chinese, English, Finnish, French, German, Italian, Japanese, Portuguese, Rumanian, Russian, Slovene and Slovak. As of now, a total of 1407 translated and glossed sentences are included in the lexical entries. The dictionary is freely available for consultation at [<https://www.uam.es/gruposinv/upstairs/v31/index.htm>].

The design of the dictionary seeks to reconcile theoretical exhaustiveness and user-friendliness by distributing different kinds of information between the three main modules (‘Minimal definition’, ‘Lexical entry’, and ‘Phraseology’) and also by offering flexible search options, which allow each user to choose the kind of data he wants to visualize[4]. The ‘Minimal definition’ module contains a basic definition of the verb and an example of its basic spatial use (see section 3.1). ‘Lexical entry’ provides a formal definition in terms of lexical structures (the argument structure, the event structure, the thematic structure, and the qualia structure) and also analyzes verbal periphrases and other sense modulations where the verb loses or changes a part of its meaning. Finally, ‘Phraseology’ registers phraseological units and syntagmatic combinations wherein each verb appears.

DICEMTO is mainly aimed at users with a linguistic background: it was conceived as a theoretical tool for linguists interested in the semantics and the syntax of motion, and in their cross-linguistic manifestations. However, some of its contents can also be dealt with by non-expert users (although no effort has been made so far to specifically adapt it to this group of users, and we are not going to elaborate on this here any further). For instance, the informal definitions in the modules ‘Minimal definition’ and ‘Phraseology’ dot not require linguistic knowledge and are illustrated with examples and glossed translations for a better understanding. One of the groups of users that might potentially benefit from consulting the dictionary are second language teachers and learners, since the latter are known to have difficulty interpreting and producing idiomatic expressions. Their meaning cannot be interpreted compositionally, and it is rarely accounted for by generic rules taught in class, in handbooks, and in learner dictionaries.

[4] A series of tutorials on the use of the dictionary are available at [<https://www.uam.es/gruposinv/appupstairs/upstairs2/tutoriales.html>; last access: July 15, 2016].

In what follows we present the first two modules of the lexical entry[5] of the verb andar 'to walk' ('Minimal definition' and 'Lexical entry')[6]. We will show how its senses can be integrated into a unique meta-entry structured in different levels of representation and how the meaning components belonging to these levels interact dynamically and generate different senses depending on the context.

3 Andar 'to walk' in DICEMTO

3.1 Minimal definition of andar

Definitions of motion verbs usually take up many lines in traditional dictionaries. Andar, for example, has twenty senses according to the Diccionario de la lengua española (DLE). The decision of including multiple senses for the same lexical item is frequently motivated by the vast combinatorial potential of these verbs. We believe that this proliferation of senses is undesirable from both the theoretical and the applied perspectives. It can be avoided or at least reduced substantially if the entries are underspecified, i.e., if they include a limited number of minimal lexical-semantic features which are constant across contexts and which determine how words can be combined (cf. Faber/Mairal 1999, Levin/Rappaport Hovav 2011, Bosque/Mairal 2012, De Miguel 2013). The assumptions underlying the GL model of the lexicon make this kind of definition possible: they license the combination of underspecified lexical features and structures by means of compositional mechanisms, which are triggered in context and yield the different interpretations of the word (cf. Pustejovsky 1995, De Miguel 2009).

As will be shown in the following sections, the definitions in DICEMTO account for the basic or central meaning of the verb and allow distinguishing the lexical entry from others at a purely linguistic level. We thus avoid one of the shortcomings of other models of definition, which quite often fail to differentiate between linguis-

[5] A terminological note is in order before we proceed. We use the term *lexical entry* to refer to:

- one of the modules of the dictionary
- the *lexicographic article*: all the information pertaining to a given verb and included in the three modules.

Unfortunately, this ambiguity in unavoidable in the context of this project.

[6] The content and structure of the module 'Phraseology' is exhaustively described in González Cobas/Serradilla Castaño (2013). De Miguel (2015b) lays out the theoretical foundations of DICEMTO in connection with the GL framework.

tically relevant and extralinguistic information, thus incurring lexicographical as well as lexicological inconsistencies[7].

We propose the following definition for andar:

(1) {Desplazarse/moverse} {a pie/dando pasos}
Lit.: 'To move {on foot/by taking steps}'

In this definition we captured the most important lexical features of andar: the motion component ('to move') and the manner-of-motion component ('on foot/by taking steps'). These will be analyzed in the following sections as a part of different lexical structures.

3.2 Lexical entry of *andar*

3.2.1 Argument Structure (AS) and Thematic Structure (TS)

Recent theoretical lexicographical projects account for the combinatorial properties of lexical items by including specifications and structures related to what is known in theoretical linguistics as subcategorization, selection or valence (see FrameNet, VerbNet and CPA, among others).

Our entries distribute this information among three structures:

a) the Argument Structure specifies the number of arguments required by the verb and their syntactic category;
b) the Thematic Structure encodes the semantic function of arguments in the predicate;
c) the Qualia Structure encodes inherent semantic features of the arguments, such as [PHYS_OBJ], [INFO], [EVENT], etc. These features are related to the properties of the real-world entities denoted by words (see section 3.2.3 for details).

The AS and TS of andar are represented in (2a) and (2b), respectively. The example in (2c) instantiates all the AS and TS components.

[7] Different perspectives on the 'linguistic vs. encyclopedic' dichotomy are discussed in Wotjak (1992), Busa et al. (2001) and Bosque (2000), among others. De Miguel (2009) presents a GL-framed approach to this issue.

(2) a. AS: N1 (por 'via' N2) (hacia 'toward'[8] /hasta /a 'to' N3) (de/desde 'from' N4) (N5) (Adv/Adj/etc.)

b. TS: AgentPath/Medium Direction/Goal Source Distance Manner

c. Juan[Agent] anduvo cojeando[Manner] tres kilómetros[Distance] por el monte[Medium] desde la Charca de la Olla[Source] al Pino Aprisquillo[Goal].
Lit: 'Juan[Agent] walked limping[Manner] three kilometers[Distance] through the hills[Medium] from Charca de la Olla[Source] to Pino Aprisquillo[Goal]'
'Juan hobbled around the hill for three kilometers, starting at Charca de la Olla and ending at Pino Aprisquillo.'

This kind of notation might seem unconventional because we do not label the prepositional phrases (in line with DDLC, for example) and single out the noun phrases instead (N1-N5 in the example above). This decision is motivated by the fact that nominal features are more relevant as far as the composition of the predicate meaning is concerned. Labels and indexes attached to the noun phrases create a cross-reference system shared by the different structures. This is one of the main properties of the lexical model we make use of and it also underlies the design of relational databases, which we adopted as the organizational model for our dictionary.

Not all the predicate members listed in the AS of motion verbs are equally important, which usually translates into different frequency counts. We do not include data on frequency in our dictionary (unlike FrameNet, among other recent projects), but we do take it into account when defining the AS. High frequency elements are almost always the "canonical" arguments, in syntactic terms. The less frequent ones are usually the optional modifiers (adjuncts), which we also include in the AS (in parentheses, to mark optionality). Very low frequency constituents are not included in the AS, for instance the beneficiary adjunct. The example in (3) is taken from ADESSE[9]:

(3) Una cosa es que le[Beneficiary] anduvieras detrás cuando eras una cría, como todas, porque era el mayor [...].
Lit.: 'One thing is that you him[Beneficiary] walked behind when you were a child, like everyone, because he was the oldest [...].'
'It is one thing for you to have been after him when you were a child, like all the other girls, because he was the oldest [...].'

[8] *Hacia* 'toward' typically introduces a Direction argument or adjunct, but here we assume that Direction is derived compositionally, as a combination of the preposition with a noun encoding the Goal.

[9] This example and all the others were translated by one of the co-authors of this paper (OB).

3.2.2 Event Structure (ES)

The notion of event structure (ES) is widely used in grammatical studies nowadays. In GL, this level of representation encodes the aspectual features of predicates in terms of subevents, temporally and hierarchically ordered parts of events. When combined, the subevents yield different types of events: states, processes, and transitions. The resulting event types are included as semantic types in lexical and conceptual ontologies (cf. Cruse 2004), and in lexicographical and computational works (for instance, in the SIMPLE ontology, cf. Lenci et al. 2000). DICEMTO adopts the event classification put forward in Fernández Lagunilla/De Miguel (1999), who derived eight event types for Spanish based on the GL event typology.

Andar denotes a process: a dynamic, durative and atelic event (see the ES representation in (4)). One of the defining features of processes is that they are homogenous: any part of the walking event has the same properties as the whole event, so that it is always true that a person walking has already walked.

(4) ES of andar: P[e1, e2, …, en]

Although andar is atelic, its beginning and endpoints can be explicitly referred to. The beginning is usually related to the Source constituent, introduced by the preposition desde in (5a) (where the reference to the starting point is reinforced by the inchoative periphrasis <echar a + infinitive> 'to start to + infinitive') and also in (5e,f). The endpoint is related to the Goal constituent, introduced by the preposition hasta 'to' (in (5d,e)) and hacia 'toward' (in (5f)). Thus, the Goal and Source constituents mark the beginning and the end of a homogeneous event without necessarily making it telic: 'to be walking from somewhere {to/toward} somewhere' does not entail '{to have reached/to have walked all the way to} somewhere'[10]. The predicate components referring to Path, Medium, Direction (in (5a,b,c,f)) and Manner (in (5g)) are associated to the main phase of the event, the process:

[10] It must be noted that the presence of *hasta* 'to' does make the event telic when it contributes to creating a complex event (i.e., a transition) composed of two phases: a process and a resultant state. This is what happens in (i), where *Juan anduvo hasta la valla* 'Juan walked to the fence' is compatible with the durative adverbial *durante dos horas* 'for two hours' (which focuses on the process subevent) and with the time-frame adverbial *en dos minutos* 'in two minutes' (which focuses on the transition as a whole).
(i) Juan anduvo hasta la valla {durante dos horas / en dos minutos}.
'Juan walked to the fence {for two hours / in two minutes}.'

(5) a. Echamos a andar desde el refugio de la Perdiz[Source] camino arriba [Direction].
'We began walking from the shelter La Perdiz[Source] uphill[Direction].'

b. ¿Qué sentirá cuando sus pies anden por ese césped[Medium] tantas veces recorido?
'What will it feel like when his feet walk on this lawn[Medium], which he went through so many times?'

c. Me encanta andar por el campo[Medium].
'I love walking in the countryside[Medium].'

d. Con los pies descalzos, anduvo hasta la orilla del mar[Goal].
'He walked to the seashore[Goal] barefoot.'

e. En cuanto se levanta y anda desde la cama[Source] hasta el sofá[Goal], ve las estrellas.
'Every time he gets up and walks from the bed[Source] to the sofa[Goal], he writhes in pain.'

f. Por la calle 155[Path] anduvimos desde Broadway[Source] hacia el este [Direction].
We walked from Broadway[Source] eastward[Direction] along the 155th Street[Path].'

g. [...] ande rápidamente[Manner] o corra algunos minutos.
'[...] walk fast[Manner] or run for a few minutes.'

In addition to the Goal constituent introduced by hasta 'to' (see footnote 10), the boundedness of the Path can be reinforced by the aspectual clitic se, which (as claimed in De Miguel/Fernández Lagunilla 2000) can only be added when the event has a culmination phase followed by a change of state, as in (6).

(6) [...] 5.600 personas los que nos anduvimos desde la Plaza de Santo Domingo hasta la Delegación del Gobierno[11].
Lit.: '(we) 5600 people walked to ourselves from Plaza de Santo Domingo to the regional Government Office.'
'There were 5600 of us that walked all the way from Plaza de Santo Domingo to the regional Government Office.'

[11] The clitic *nos* in (6) has a benefactive meaning in addition to the aspectual delimitative value. It might very well be possible that this is an evaluative benefactive clitic rather than an aspectual clitic, as argued in Armstrong (2013). Be it as it may, it does exert an impact on the aspectual makeup of the predicate.

3.2.3 Qualia Structure (QS)

The qualia structure (QS) is the most ground-breaking and also the most controversial component of the GL theory. It is novel because it extends the domain of linguistically relevant information to semantic features usually considered as extralinguistic, i.e., pertaining to real-world knowledge rather than to linguistic knowledge.

In GL, these features are encoded in the lexical entry whenever they are linguistically relevant, i.e., if they can be shown to determine the syntactic behavior of the word. For instance, the [±dynamic] feature allows explaining the different meanings of the verb llegar 'to come, to reach' in (7) (examples taken from De Miguel 2004b):

(7) a. El atleta llega a la meta.
'The athlete is reaching the finishing line.'
b. El niño llega al botón del ascensor.
'The boy {reaches / is reaching} the lift button.'
c. La carretera llega hasta el valle.
'The road goes all the way to the valley.'

When combined with atleta 'athlete', which is [+dynamic], llegar denotes a motion event but with carretera 'road', which is [-dynamic], it does not: since carretera refers to a static object, which cannot move, the event loses the dynamic motion phase and turns into a mere state (of the road having a certain extension). (7b) is ambiguous between a dynamic and a static interpretation because el niño 'the boy' can be conceptualized as a moving dynamic entity or as a static entity endowed with spatial extension. In the former case el botón del ascensor 'the lift button' is the Goal of a motion event and in the latter it is the upper limit of the subject's vertical extension.

The [±dynamic] feature crucially affects the syntactic behaviour of the predicate: the [+dynamic] predicate in (8a,a') is compatible with the progressive form and rejects the adverbial desde hace años 'for years', and the [-dynamic] predicate in (8b, b') rejects the former and accepts the latter:

(8) a. El atleta está llegando a la meta en este momento.
‘The athlete is reaching the finishing line at this moment.’

a’. El atleta llega a la meta desde hace años.
Lit.: ‘The athlete reaches the finishing line for years.’

b. *La carretera está llegando hasta el valle en este momento.
The road is reaching the valley at this moment.’

b’. La carretera llega hasta el valle desde hace años.
Lit.: ‘The road reaches the valley for years.’
‘The road reached the valley years ago.’

Since the [±dynamic] feature determines the syntactic behavior of the verbal predicates, it must be included in the QS of atleta and carretera[12]. However, this is only one part of the story. The new meaning of the verb cannot be derived compositionally if it is not encoded in the lexicon, and one of the basic premises of DICEMTO is that the underspecified verbal definition must accommodate it.

The qualia roles refer to the meaning parameters that define events, objects and properties from the point of view of their origin (agentive role); their physical characteristics (size, shape, etc.) and ontological classification (formal role); their internal constitution (constitutive role); and their purpose (telic role). As pointed out above, their appropriateness in a grammatical analysis might seem debatable (cf. De Miguel 2009 and the references therein). In lexicography, on the other hand, their relevance is generally taken for granted. Thus, the Spanish lexicographer Julio Casares makes use in his renowned Introducción a la lexicografía moderna (1950) of the real definitions, which “aim at discovering the nature, the essence of the denoted thing”, nominal definitions, which “merely explain the word meaning”, genetic definitions, which “account for the defined concept as an effect of the action”, teleological definitions, which “inform us about the nature of things as determined by their function”, and descriptive definitions, which “combine the description of the form and other relevant properties with teleological or functional specifications, and also with the cause or origin of the defined thing”. There is an obvious similarity between the genetic definition and the values encoded in the agentive role of the QS, between the teleological definition and the telic role, and between the descriptive definition and the formal role (cf. Batiukova 2009a, §3.1.).

It seems clear that the qualia are very well suited to describe nouns, especially the concrete ones. The formal role values are similar to the semantic selection features of the generative grammar and to the top-level nodes of lexical-conceptual

[12] The semantics of motion verbs is discussed in De Miguel (2004b, 2012) and Batiukova (2008: ch. 4-5), and the syntactic consequences of certain QS values are deal with in De Miguel (2004a, 2009, 2011, 2015a).

ontologies (e.g., WordNet). In the following definition of cocoa, taken from the Merriam-Webster Online Dictionary[13], we can easily detect pieces of information corresponding to the four qualia:

cocoa *noun*

a. [powdered ground roasted]$_{\text{Agentive/Constitutive}}$ [cacao beans]$_{\text{Formal}}$ from which a portion of the [fat]$_{\text{Constitutive}}$ has been removed and which [is used to make chocolate]$_{\text{Telic}}$

b. a [beverage]$_{\text{Formal}}$ [prepared by heating cocoa with water or milk]$_{\text{Agentive/Constitutive}}$

The lexical entries of DICEMTO specify the formal role values of all the nouns included in the AS and the ES of the verbs.

The Agent argument (N1) of andar is a dynamic entity that covers a certain distance (N5) and follows a path (N2, a location with spatial extension) between two locations: the Source (N4) and the Goal (N3). The adverbial or adjectival component encoding manner of motion is associated to the syntactic subject through its agentive role, because the Agent of andar is usually the internal dynamic cause of the motion event (cf. Levin/Rappaport Hovav 1992, Pustejovsky/Busa 1995, and Batiukova 2008):

(9) QS of *andar*
N1: [F = dynamic entity]
N2: [F = location∧spatial extension]
N3: [F = location]
N4: [F = location]
N5: [F = linear measure]
Adv/adj/etc.: [A = manner of the Agent or dynamic entity]

The semantic features represented in (9) are the canonical ones, they define the spatial dynamic uses of andar. However, it quite often exhibits apparently non-literal meanings when combined with arguments whose semantic type is different from the ones stipulated in the canonical QS. For instance, (10a) encodes a spatial non-dynamic event and (10b) encodes a non-spatial dynamic event.

(10) a. El libro que busca Juan anda por algún lugar de tu casa.
Lit.: 'The book that Juan is looking for walks around some place in your house.'
'The book Juan is looking for is somewhere in your house.'

[13] [<https://www.merriam-webster.com/>; last access: July 9, 2016].

b. Juan anda preguntando desde hace días por su libro.
Lit.: 'Juan walks around asking for days for his book.'
'Juan has been asking for his book for days.'

The challenge consists in integrating these cases into our lexical entry and accounting for the semantic and grammatical alterations yielded by the combination of andar with different semantic types of arguments.

3.3 Modifications: the canonical definition of *andar* and its alterations

A quick look at the Word Sketch of andar[14] suffices to realize that, on the one hand, the dynamic entities are but a small subset of the syntactic subjects of this verb and, on the other hand, that in many cases the verb loses its spatial meaning.

Table (11) classifies the uses of andar based on the QS of its subjects. The first two columns list the subjects of the verb and their absolute frequency. The next five columns show the interpretation that the verbal predicate acquires when combined with each subject in terms of the possible combinations of the features [±spatial] and [±dynamic].

(11) Andar, esTenTen11 (European, TreeTagger)

Head N of the subject	**Overall frequency**	**[+spatial +dynamic]**	**[+spatial -dynamic]**	**[-spatial, +dynamic]**		**[-spatial -dynamic]**
				"Metaphoric" and periphrastic uses	**andar +{adv/adj/PP/participle}**	
cosa 'thing'	236		X	X	X	
gente 'people'	128	X	X	X	X	
chico 'boy, guy'	92	X		X	X	
cojos 'lame people'	32	X				

[14] Word Sketch is an automatically generated summary of the sentential elements that a given lexical item is combined with in a corpus (Kilgarriff et al. 2008). We used the esTenTen web corpus integrated into the Sketch Engine corpus query system.

Head N of the subject	**Overall frequency**	**[+spatial +dynamic]**	**[+spatial -dynamic]**	**[-spatial, +dynamic]**		**[-spatial -dynamic]**
				“Metaphoric” and periphrastic uses	**andar +{adv/adj/ PP/participle}**	
PP ‘PP (Popular Party)’	37			X	X	
político ‘politician’	24	X		X	X	
moto ‘motorcycle’	20	X		X	X	
Barça [a football club]	18			X	X	
diablo ‘devil’	17				X	
burro ‘donkey’	12	X				
Beti[s] [a football club]	11			X	X	
Ferrari	10	X		X	X	
francés ‘French’	9		X	X	X	
muchacho ‘kid, boy’	9			X	X	
chaval ‘kid, buddy’	9	X		X	X	
gallina ‘hen’	8	X		X	X	
peña ‘group, club’	8	X		X	X	
torero ‘bullfighter’	8			X	X	
militar ‘military man’	8	X		X	X	
republicano ‘republican’	8		X	X	X	
fantasma ‘ghost’	6	X		X	X	
hyenas ‘hyenas’	6			X	X	

Head N of the subject	**Overall frequency**	**[+spatial +dynamic]**	**[+spatial -dynamic]**	**[-spatial, +dynamic]**		**[-spatial -dynamic]**
				“Metaphoric” and periphrastic uses	**andar +{adv/adj/PP/participle}**	
Ibex	6			X	X	
cazador ‘hunter’	6			X	X	
paralíticos ‘paralytic (pl.)’	5	X				
lobo ‘wolf’	5		X	X	X	
ladrón ‘thief’	5	X	X	X		
Sporting [a football club]	5				X	
economía ‘economy’	5			X	X	
CIA	5			X	X	
Vaticano ‘the Vatican’	5			X	X	
sevillano ‘Sevillian’	5				X	
vasco ‘Basque’	4	X		X	X	
andaluz ‘Andalusian’	4	X			X	
chorizo ‘thief’	4			X	X	
parejita ‘couple’	3	X			X	
Cid	3	X			X	
Juventus [a football club]	3			X	X	
granadinos ‘of Granada (pl.)’	3			X	X	

Head N of the subject	Overall frequency	[+spatial +dynamic]	[+spatial -dynamic]	[-spatial, +dynamic]		[-spatial -dynamic]
				“Metaphoric” and periphrastic uses	andar +{adv/adj/PP/participle}	
demonios ‘demons’	3				X	
Cope [radio station]	3				X	
maya ‘Maya/Mayan’	3				X	
FARC ‘Revolutionary Armed Forces of Colombia’	3				X	
Bayern [a football club]	3				X	
duende ‘goblin’	3				X	
Arsenal [a football club]	3				X	
parapléjicos ‘paraplegic (pl.)’	2	X				
Mossad	2				X	
pendejos ‘nerds’	2				X	
pederasta ‘pedophile’	2				X	

3.3.1 *Andar* in [+spatial, +dynamic] contexts

Most dynamic entities that participate in [+spatial, +dynamic] events headed by andar are represented by animate human nouns, both individual and collective (chico ‘boy, guy’, chaval ‘kid, buddy’, parejita ‘couple’, republicano ‘republican’, político ‘politician’, militar ‘military man’, vasco ‘Basque’, andaluz ‘Andalusian’, Cid, paralíticos ‘paralytic (pl.)’, parapléjicos ‘paraplegic (pl.)’, gente ‘people’, peña 'group, club'), names of animals (burro 'donkey', gallina 'hen'), and also names of artifacts designed to generate motion (moto 'motorcycle', Ferrari).

As mentioned in section 3.2.3, the manner-of-motion component of the predicates headed by andar is encoded in the agentive role of the Agent subject. This component can refer to either the manner in which the event unfolds or the state of the subject while performing the action, and it can be expressed as an adverb, an adjective, a participle or a prepositional phrase. In addition, as all manner-of-motion verbs, andar lexicalizes a specific way of moving, which we captured as 'on foot/by taking steps' in the definition (cf. section 3.1).

Studies framed within the Generative Lexicon model (Pustejovsky/Busa 1995, Batiukova 2008) have proposed that the manner-of-motion component is encoded in the formal role of the QS of the verb and that this feature must be compatible with the information encoded in the formal role of the noun referring to the moving Agent. The feature 'on foot/by taking steps' is compatible with the meaning of human nouns and some names of animals, because these have extremities that end in feet or paws and therefore are able to move 'on foot/by taking steps': chico 'boy, guy', chaval 'kid, buddy', parejita 'couple', republicano 'republican', político 'politician', militar 'military man', vasco 'Basque', andaluz 'Andalusian', burro 'donkey', gallina 'hen', etc.

The same analysis cannot be applied to Ferrari or moto 'motorcycle' because the features encoded in the formal role of the verb and of these two nouns apparently do not match. Even so, we can easily find expressions such as (12a) and (12b) in the corpus, and the speakers have no difficulty interpreting them.

(12) a. El Ferrari anduvo como una bala.
Lit.: 'The Ferrari walked like a bullet.'
'The Ferrari was as fast as a bullet.'

b. La moto anda menos.
Lit.: 'The motorcycle walks less.'
'The motorcycle is going slower.'

We believe that these expressions are acceptable because the meaning of andar undergoes an adjustment process called co-composition (Pustejovsky 1995, De Miguel/Batiukova 2017), whereby the argument imposes its selectional requirements on the verb. In this particular case, the manner of motion expressed by the predicate as a whole is imposed by the lexical entry of the Agent instead of the verb: a Ferrari or a motorcycle move when their wheels spin, and not 'by taking steps'.

3.3.2 Andar in [+spatial, -dynamic] contexts

The meaning alterations discussed in the following three sections differ to a lesser or greater extent from the basic motion meaning encoded in the minimal definition ‘to move on foot/by taking steps’.

As shown in Table (11), the motion component is absent in many of the retrieved sentences. In some cases the event keeps being spatial but becomes non-dynamic, as in (13), where andar means ‘be located somewhere’:

(13) a. El sol se ha puesto, y los lobos *andan* cerca. (=‘están cerca’)
Lit.: ‘The sun set and the wolves *walk* nearby.’ (=‘are nearby’)

b. Sabemos que los ladrones *andan* en los estacionamientos, esperando que el dueño se aleje de su vehículo. (=‘están en los estacionamientos’)
Lit.: ‘We know that the thieves *walk* around the parking lots, waiting for the car owner to leave his vehicle.’ (=‘are in the parking lots’)

The events in (13) are stative. If we change the verbal form for progressive, the expression remains acceptable but acquires a spatial dynamic interpretation:

(14) a. Los lobos *están andando* cerca.
Lit.: ‘The wolves *are walking* nearby.’(=‘are prowling nearby’)

b. Los ladrones *están andando* en los estacionamientos.
Lit.: ‘The thieves *are walking* around the parking lots.’ (=‘are prowling around the parking lots’)

De Miguel (2012) argues that the static uses of andar can be accounted for if we assume that any event can be decomposed into a sequence of states and transitions between these states (this claim is implicit in Pustejovsky 1991 and 1995, and it is further elaborated in Moreno Cabrera 2003 and Mani/Pustejovsky 2012, among others). When the [+dynamic] feature is lost, as in (13), and there is no transition from one state to another, the process is reduced to a state and the Path constituent comes to denote the spatial location of the subject.

However, even in these static uses of andar there is a residue of dynamicity, which is why it is not completely synonymous with the spatial meaning of estar ‘be located somewhere’. The predicates with andar imply a sequence of spatial or temporal states, as in (15). In (15a), en Lima ‘in Lima’ is the spatial frame wherein the iteration of states takes place (meaning roughly ‘happening {over there / over and over again}’). In (15b), the spatial frame within which multiple states of ‘being in the Internet’ occur is por ahí ‘over there / around’.

(15) a. Las cosas en Lima andan complicadas y debo regresar allá.
Lit.: 'The things in Lima walk complicated and I must go back there.'
'The things in Lima have been complicated and I must go back there.'

b. [...] desde luego no es normal que la gente ande por ahí navegando con un navegador de hace 20 años.
Lit.: '[...] of course it is not normal that the people walk navigating over there with a 20-year-old browser.'
'Of course it is not normal that some people are using a 20-year-old browser to navigate.'

The examples in (14) and (15) illustrate the possibility of splitting the features [±spatial] and [±dynamic] in the meaning of andar. When both are positively valued, the predicate has the basic motion meaning. When there is no dynamicity but the spatial feature is preserved, a stative locative meaning emerges, as in (14). The iteration of a state or location sometimes reveals a residue of dynamicity (as argued above for the sentences in (15)), which is absent in the spatial use of estar 'be located somewhere'.

As will be shown right away, we can also focus on the change of state in time instead of space, and derive predicates with andar defined as [-spatial, +dynamic].

3.3.3 *Andar* in [-spatial, +dynamic] contexts

3.3.3.1 Andar denoting motion in time

In the cases discussed in this section, the predicates with andar are dynamic and they present their arguments syntactically as a moving Agent, Goal or Path. However, they do not express motion in space. This section focuses on constructions with adjectives, adverbs, participles and prepositional phrases (as in (16))[15], and the next section discusses verbal periphrases with andar.

In (16a), the Agent subject los granadinos Eskorzo 'the band from Granada Eskorzo' cannot physically move along the Path denoted by the constituent por el camino de la evolución musical 'down the road of musical evolution'. In (16b), the Agent los cazadores 'the hunters' cannot literally go after the prey because the noun

[15] Following Camus Bergareche (2006), we assume that these constructions are attributive, since the verb merely contributes the continuative or iterative meaning to the state description encoded by the adjective, adverb, participle or PP. For this reason, the sentence in (i) can be paraphrased as 'Juan has been concerned about his health', with the copula estar in Spanish:
(i) Juan anda preocupado por su salud. (='Juan está preocupado por su salud')
Lit.: 'Juan walks concerned about his health.'
'Juan has been concerned about his health.'

liebre ‘hare’ refers to an unexpected event (or previously unknown information) rather than to a moving animal (ha saltado la liebre means ‘something unexpected happened’). The subjects in (16c) and (16d) show the same behavior: la economía ‘economy’ and la cosa ‘the thing’ are not capable of moving, and there are no other elements in the predicate that can yield a different interpretation (like Lima or por ahí ‘over there’ in (15)). One of the possibilities that is left is to switch the interpretation from the spatial domain to the temporal domain: the state denoted by the predicate is iterated in time (with more or less significant changes) either explicitly (hace trece años ‘for thirteen years’ in (16a), poco a poco ‘little by little’ in (16b)) or implicitly, as in (16c) and (16d), which are compatible with the specification ‘lately / over and over again’ and ‘back then, during that period’, respectively. The verbal tense – present simple in (16c) and simple perfect in (16d)– renders this interpretation possible[16].

(16) a. Hace ya casi trece años que los granadinos Eskorzo[moving Agent] andan constantemente por el camino de la evolución musical[Path].
Lit.: ‘The Eskorzo band from Granada have been constantly walking down the road of musical evolution for thirteen years.’

b. La liebre ha saltado en el Ayuntamiento de Murcia, y los cazadores[moving Agent] andan tras su presa[Goal], a la que van cercando poco a poco y en lógica capturarán.
Lit.: ‘The hare has jumped in the city hall of Murcia and the hunters walk after their prey, which they are surrounding little by little and which they will most certainly capture.’

c. La economía anda muy mal.
Lit.: ‘The economy walks very badly.’
‘The economy is in bad shape.’

d.. […] la cosa anduvo más repartida.
Lit.: ‘[…] the thing walked more (evenly) distributed.’
‘Things used to be distributed more evenly.’

[16] Here we deal with the same iterative meaning of the simple perfect as in (i) and (ii). One of the sources of the iterative interpretation of (i) is the non-referential nature of the determinerless direct object *coche* ‘car’: the speaker always had an object of the kind ‘car’, but it most likely was a different instance of this object (a different car) on each occasion, which renders the state of ‘having a car’ iterative.

(i) Siempre tuvo coche.
‘He always had a car / He used to always have a car.’

(ii) Nunca estuvo enfermo.
‘He never got seek.’

Examples like (16) are usually analyzed as metaphorical, but we believe them to be the outcome of the lexical agreement processes between the minimal definition of the verb and the QS features of its nominal arguments.

3.3.3.2 Verbal periphrases with *andar*

Andar also loses its spatial meaning and retains the [+dynamic] feature in the aspectual periphrasis <andar + gerund>. The periphrasis can usually be replaced with the main verb but there are some subtle aspectual differences.

Verbs appearing in periphrastic constructions as auxiliaries undergo semantic bleaching: they lose their lexical content to a lesser or greater extent, and merely express very general grammatical meanings (temporal, aspectual and modal). Yet the bleaching is never complete: the auxiliary usually retains a part of the meaning encoded by the verb in its lexical uses. The more general and polysemic the verb meaning is, the higher its capacity of losing lexical content. Venir 'to come', for instance, participates as auxiliary in four periphrases (see Batiukova/De Miguel 2013a). By contrast, rich and specific semantic content correlates with inability to form periphrases. This would be the case of volver 'to return, to come back' and andar 'to walk': the former lexicalizes a very specific kind of path, the latter encodes a peculiar manner of motion, and they only form one periphrasis each (<volver a + infinitive> and <andar + gerund>).

This part of the lexical entry of andar is based on the description in Martínez-Atienza (2006) and Nueva gramática de la lengua española (2009), which we redefined in terms of the event structure typology put forward in De Miguel/Fernández Lagunilla (2000, 2004, 2007).

We argue that the periphrasis <andar + gerund> has a continuative meaning, similar to <estar + gerund>: it focuses on the intermediate phase of the event without referring to its endpoint. In other words, this periphrasis encodes processes (P) and it has the same event type as andar. However, unlike andar, <andar + gerund> can express frequentative meaning when it denotes situations that unfold with interruptions or intermittently (we will elaborate on this when analyzing the main verbs compatible with this periphrasis).

Verbal periphrases generally impose aspectual constraints on their main verb, and our dictionary accounts for these constraints[17]. Since <andar + gerund> encodes processes, it is compatible with durative events not oriented towards a telos (see Martínez-Atienza 2006: 87): processes (as in (17a,b)), iterative achievements (17c,d), iterative accomplishments (17e), detelicized accomplishments (17f), and non-permanent states (17g). Permanent states are usually incompatible with continua-

[17] For simplicity's sake, we use the traditional Vendlerian terminology here instead of the more structured proposals by Pustejovsky (1995) and De Miguel/Fernández Lagunilla (2000, 2004, 2007). The latter are used in DICEMTO to define the overall event type expressed by the periphrasis.

tive and progressive periphrases because these involve dynamicity which the permanent states lack: their course is independent from the temporal variable.

(17) a. Pero cómo te vas a ir si es ya de noche y los chacales y las hienas andan merodeando por ahí. [PROCESS]
'You cannot possibly leave now: it is dark already, and the jackals and hyenas are walking and lurking around.' / 'You cannot possibly leave now: it is dark already, and the jackals and hyenas are lurking around.'

b. [...] el Ibex andaba luchando por no perder los 9.800 puntos. [PROCESS]
'Ibex has been struggling to hold on to the 9,800-point level.' [lit.: 'walked struggling']

c. Estos días el PP anda proclamando que pueden bajar impuestos a muerte y a saco. [ITERATIVE ACHIEVEMENT]
'The Popular Party has been claiming these days that they are determined to cut taxes across the board.' [lit.: 'walks proclaiming that']

d.. La gente andaba golpeando un bloque de hielo. [ITERATIVE ACHIEVEMENT]
'The people were pounding at an iceblock.' [lit.: 'walked pounding']

e. Ya sé que con cuarenta y tres años soy algo mayor ya para andar escribiendo cartas de estas. [ITERATIVE ACCOMPLISHMENT]
'I know that at forty-three years old I'm a little too old for writing this kind of letters.' [lit.: 'walk writing']

f. [...] de momento ando esculpiendo mi nueva criatura [DETELICIZED ACCOMPLISHMENT]
'[...] for now I'm working on my new sculpture.' [lit.: 'walk sculpting my new creature']

g. Las chicas andan disfrutando de un baño de burbujas. [NON-PERMANENT STATE]
'The girls are enjoying a bubble bath.' [lit.: 'walk enjoying']

In (17b-g) andar heads non-spatial predicates. (17a) can have three different interpretations: (a) andar as a lexical verb ([+spatial, +dynamic]) with predicative gerund ('the jackals and hyenas walk around and lurk'); (b) andar as an attributive verb (close in meaning to the spatial non-dynamic use of estar 'be located') with an adverbial gerund; (c) andar as auxiliary of continuative periphrasis, i.e., [-spatial, +dynamic], as in all the other predicates in (17).

The iterative examples preserve the dynamicity component lexicalized by andar: the examples (17b,c) can be paraphrased as 'to do something over and over again'. Andar also seems to be determining the continuative-intermittent meaning of (17a,b,d) ('to be lurking intermittently, while doing other things', 'engage in several episodes of struggle', 'be sculpting something in parallel with other activities'): since the main verbs merodear 'lurk, prowl', luchar 'struggle' and esculpir 'sculpt' do not denote intermittent events, the continuative-intermittent meaning must be contributed by andar, whose residual dynamicity emphasizes the transitions between the states within the macroevent and thus favors a fragmented vision thereof. To sum up, the main contribution of andar to the meaning of the sentences in (17) consists in focusing on the subject's passing through the different temporal states that make up the event denoted by the predicate.

3.3.4 Andar in [-spatial, -dynamic] contexts

As shown in (11), we detected no non-spatial and non-dynamic predicates. Section 3.3.2 discussed examples (recall (15)) that can in principle be paraphrased with a stative verb but still retain a residue of dynamicity.

Truly stative non-spatial cases are very rare, the sentence in (18) being one of them: here andar means 'be functional' when applied to an artifact (a watch). One of the sentential elements contributing to the stative interpretation is the negation.

(18) El reloj no anda.
Lit.: 'The watch does not walk.'
'The watch is broken.'

This seems to prove that the original meaning of the lexical verb cannot be totally cancelled no matter what context it is surrounded by. In the case of andar, the meaning component that persists even in the non-spatial uses is dynamicity.

4 Conclusion

This paper presented the research project "Multilingual Electronic Dictionary of Motion Verbs". We have outlined its theoretical foundations and the basic tenets of the analysis we have been applying to both spatial and non-spatial uses of motion verbs, including the so-called metaphorical extensions and idiomatic expressions. We believe to have proven that this kind of research can make important contributions to the theoretical study of motion verbs and to their lexicographic representation.

5 Literature

5.1 Dictionaries and lexical databases

ADESSE = Base de datos de Verbos, Alternancias de Diátesis y Esquemas Sintáctico-Semánticos del Español. Dir. García-Miguel, José María. Vigo: Universidad de Vigo. [<http://adesse.uvigo.es/>].

CPA = Corpus Pattern Analysis. Dir. Hanks, Patrick. Brno: Masaryk University. [<http://nlp.fi.muni.cz/projects/cpa/>].

DDLC = Diccionari Descriptiu de la Llengua Catalana. Barcelona: Institut d'Estudis Catalans. [<http://dcc.iec.cat/ddlc/index.asp>].

DICEMTO = Diccionario electrónico multilingüe de verbos de movimiento. Dir. De Miguel, Elena. Madrid: Universidad Autónoma de Madrid. [<https: //www.uam.es/gruposinv/upstairs/upstairs2/ agentivo.html>].

DLE = Diccionario de la lengua española. Real Academia Española. Madrid: Espasa, 23rd ed., 2014.

REDES = Diccionario combinatorio del español contemporáneo. Dir. Bosque, Ignacio. Madrid: SM, 2004.

Spanish FrameNet. Dir. Subirats, Carlos. Universidad Autónoma de Barcelona / International Computer Science Institute (Berkeley, CA). [<http://gemini.uab.es:9080/SFNsite>].

Sketch Engine. Dir. Kilgarriff, Adam. Lexical Computing Ltd. [<www.sketchengine.co.uk>].

VerbNet. Dir. Palmer, Martha. University of Colorado. [<https://verbs.colorado.edu/~mpalmer/projects/verbnet.html>].

WordNet. Dir. Fellbaum, Christiane. Princeton University. [<http://wordnet.princeton.edu/>].

5.2 Monographes and articles

Armstrong, Grant (2013): The Evaluative Dative Reflexive in Spanish. In: Borealis 2 (2), 81–128.

Batiukova, Olga (2008): Del léxico a la sintaxis: aspecto y qualia en la gramática del ruso y del español (CD-ROM). Madrid: Ediciones de la Universidad Autónoma de Madrid.

Batiukova, Olga (2009a): Aplicaciones lexicográficas de la teoría del Lexicón Generativo. In: De Miguel, Elena et al. (eds.): Fronteras de un diccionario. Las palabras en movimiento. San Millán de la Cogolla: Cilengua, 231–268.

Batiukova, Olga (2009b): La teoría del léxico en los nuevos diccionarios. In: De Miguel, Elena (ed.): Panorama de la Lexicología. Barcelona: Ariel, 487–519.

Batiukova, Olga/De Miguel, Elena (2013): Tratamiento lexicográfico de verbos de movimiento con significado amplio. In: Cabedo Nebot, Adrián et al. (eds.): Estudios de lingüística: investigaciones, propuestas y aplicaciones. Valencia: Tecnolingüística, 439–449.

Bosque, Ignacio (2000): Objetos que esconden acciones. Una reflexión sobre la sincategorematicidad. In: Cabré, María Teresa/Gelpí, Cristina (eds.): Lèxic, corpus i diccionaris, Cicle de conferències i seminaris 97-98. Barcelona: Universitat Pompeu Fabra, IULA, 15–30.

Bosque, Ignacio/Mairal, Ricardo (2012): Definiciones mínimas. In: Rodríguez González, Félix (ed.): Estudios de Lingüística Española. Homenaje a Manuel Seco. Alicante: Publicaciones de la Universidad de Alicante, 119–132.

Busa, Federica et al. (2001): Generative Lexicon and the SIMPLE Model: Developing Semantic Resources for NLP. In: Bouillon, Pierrette/Busa, Federica (eds.): The Language of Word Meaning. New York: Cambridge University Press, 333–349.

Camus Bergareche, Bruno (2006): Andar + participio. In: García Fernández, Luis (dir.): Diccionario de perífrasis verbales. Madrid: Gredos, 90–91.

Casares, Julio (1992) [1950]: Introducción a la lexicografía moderna. 3rd ed. Madrid: CSIC.

Cruse, Alan (2004): Meaning in Language. Oxford/New York: Oxford University Press.

De Miguel, Elena (2004a): La formación de pasivas en español. Análisis en términos de la estructura de qualia y la estructura eventiva. In: Verba Hispanica XII, 107–129.

De Miguel, Elena (2004b): Qué significan aspectualmente algunos verbos y qué pueden llegar a significar. In: Cifuentes, José Luis/Marimón, Carmen (coords.): Estudios de Lingüística: el verbo. Alicante: Estudios de Lingüística de la Universidad de Alicante, 167–206.

De Miguel, Elena (2009): La Teoría del Lexicón Generativo. In: De Miguel, Elena (ed.): Panorama de la Lexicología. Barcelona: Ariel, 337–368.

De Miguel, Elena (2011): En qué consiste ser verbo de apoyo. In: Escandell, Victoria/Leonetti, Manuel/Sánchez, Cristina (eds.): 60 Problemas de Gramática (dedicados a Ignacio Bosque). Madrid: Akal, 139–146.

De Miguel, Elena (2012): Verbos de movimiento en predicaciones sin desplazamiento espacial. In: Kalenić Ramšak, Branka et al. (eds.): Actas del III Simposio Internacional "La percepción del tiempo en lengua y literatura" (Ljubljana, November 24-26, 2011). Monographic volume of Verba Hispanica XX (1), 185–210.

De Miguel, Elena (2013): La polisemia de los verbos soporte. Propuesta de definición mínima. In: Torner Castells, Sergi/Bernal Gallén, Elisenda (eds.): Los verbos en el diccionario. In: Anexos Revista de Lexicografía 20, 67–109.

De Miguel, Elena (2015a): Los nombres psicológicos: propuesta de análisis en términos sub-léxicos. In: Marín, Rafael (ed.): Los predicados psicológicos. Madrid: Visor, 211–248.

De Miguel, Elena (2015b): Minimal Definitions and Lexical Agreement: Project of a Dynamic Dictionary. In: Silvestre, João Paulo/Villalva, Alina (eds.): Planning non-existent dictionaries. Lisboa: Centro de Linguística da Universidade de Lisboa/Universidade de Aveiro, 69–102.

De Miguel, Elena/Batiukova, Olga (2017): Compositional mechanisms in a generative model of the lexicon. In: Torner Castells, Sergi/Bernal Gallén, Elisenda (eds.): Collocations and other lexical combinations in Spanish. Theoretical, Lexicographical and Applied Perspectives, London/New York: Routledge, 92-113.

De Miguel, Elena/Fernández Lagunilla, Marina (2000): El operador aspectual se. In: Revista Española de Lingüística 30 (1), 13–43.

De Miguel, Elena/Fernández Lagunilla, Marina (2004): Un enfoque subeventivo de la relación entre predicados secundarios y adverbios de manera. In: Revue Romane 39 (1), 22–44.

De Miguel, Elena/Fernández Lagunilla, Marina (2007): Sobre la naturaleza léxica del aspecto composicional. In: Actas del VI Congreso de Lingüística General. Madrid: Arco/Libros, 1767–1778.

Faber/Mairal, Ricardo (1999): Constructing a Lexicon of English Verbs. Berlin/New York: De Gruyter.

Fernández Lagunilla, Marina/De Miguel, Elena (1999): Relaciones entre el léxico y la sintaxis: operadores de foco y delimitadores aspectuales. In: Verba 26, 97–128.

González Cobas, Jacinto/Serradilla Castaño, Ana (2013): Unidades fraseológicas con verbos de movimiento. Propuestas para un diccionario. In: Círculo de Lingüística Aplicada a la Comunicación 54, 7–43.

Kilgarriff, Adam et al. (2008): The Sketch Engine. In: Fontenelle, Thierry (ed.): Practical Lexicography. A reader. Oxford, New York: Oxford University Press, 297–306.

Lenci, Alessandro et al. (2000): SIMPLE Linguistic Specifications. Deliverable 2.1. Pisa: University of Pisa/Institute of Computational Linguistics of CNR. [http://www.ilc.cnr.it/AZ_bibliography/Z176.PDF, last access: July 10, 2017].

Levin, Beth/Rappaport Hovav, Malka (1992): The lexical semantics of verbs of motion: the perspective from unaccusativity. In Roca, Iggy (ed.): Thematic Structure: Its Role in Grammar. Berlin/New York: De Gruyter, 247–270.

Levin, Beth/Rappaport Hovav, Malka (2011): Conceptual categories and linguistic categories. Course handouts, LSA Linguistic Institute. Boulder: University of Colorado.

Mani, Inderjeet/Pustejovsky, James (2012): Interpreting Motion. Grounded Representations for Spatial Language. Oxford: Oxford University Press.

Martínez-Atienza, María (2006): Andar + gerundio. In: García Fernández, Luis (dir.): Diccionario de perífrasis verbales. Madrid: Gredos, 85–90.

Moreno Cabrera, Juan Carlos (2003): Semántica y gramática. Sucesos, papeles semánticos y relaciones sintácticas. Madrid: Antonio Machado Libros.

Pustejovsky, James (1991): The syntax of event structure. In: Levin, Beth/Pinker, Steven (eds.): Lexical and Conceptual Semantics. Cambridge/Oxford: Blackwell, 47–81.

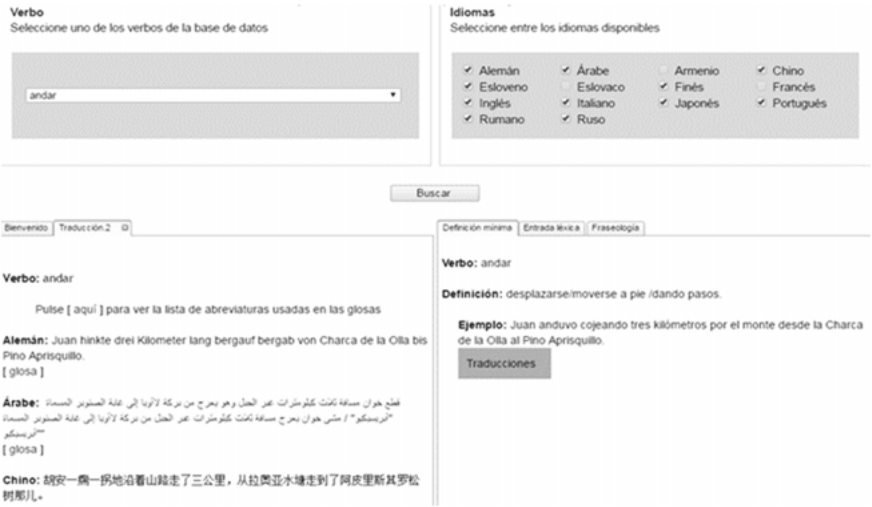
Pustejovsky, James (1995): The Generative Lexicon. Cambridge, Massachusetts: The MIT Press.

Pustejovsky, James/Busa, Federica (1995): Unaccusativity and Event Composition. In: Bertinetto, Pier Marco (ed.): Temporal Reference, Aspect and Actionality. Torino: Rosenberg, Sellier, 159–177.

Real Academia Española/Asociación de Academias de la Lengua Española (2009): Nueva gramática de la lengua española. Madrid: Espasa.

Wotjak, Gerd (1992): Estructuras en el léxico. In: Wotjak, Gerd (ed.): Estudios de lexicología y metalexicografía del español actual. Tübingen: Max Niemeyer, 108–124.

6 Appendix



Verbo
Seleccione uno de los verbos de la base de datos

andar ▾

Idiomas
Seleccione entre los idiomas disponibles

☑ Alemán	☑ Árabe	☐ Armenio	☑ Chino
☑ Esloveno	☐ Eslovaco	☑ Finés	☐ Francés
☑ Inglés	☑ Italiano	☑ Japonés	☑ Portugués
☑ Rumano	☑ Ruso		

Verbo: andar

Pulse [aquí] para ver la lista de abreviaturas usadas en las glosas

Alemán: Juan hinkte drei Kilometer lang bergauf bergab von Charca de la Olla bis Pino Aprisquillo.
[glosa]

Árabe: قطع خوان مسافة ثلاث كيلومترات عبر الجبل وهو يعرج من بركة لاأويا إلى غابة الصنوبر المسماة "أبريسكيو" / مشى خوان يعرج مسافة ثلاث كيلومترات عبر الجبل من بركة لاأويا إلى غابة الصنوبر المسماة "أبريسكيو"
[glosa]

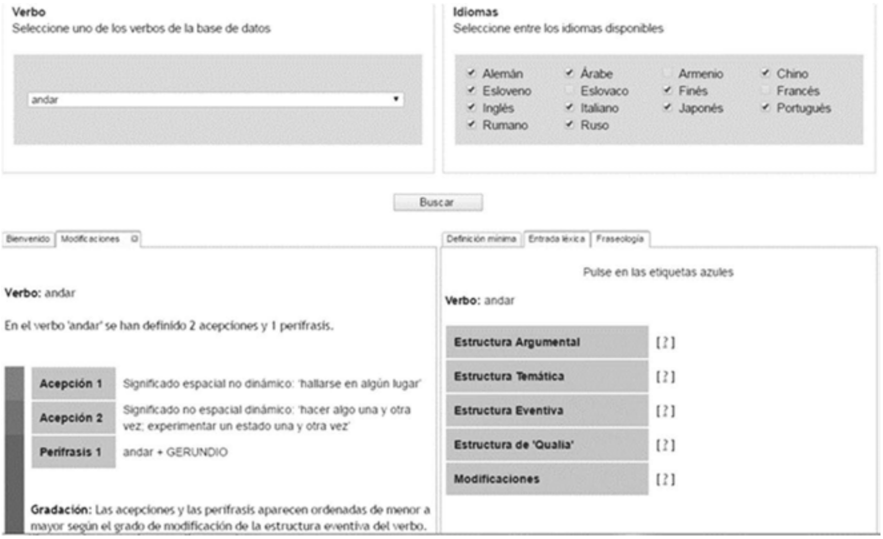
Chino: 胡安一瘸一拐地沿着山路走了三公里，从拉奥亚水塘走到了阿皮里斯其罗松树那儿。

Verbo: andar

Definición: desplazarse/moverse a pie /dando pasos.

Ejemplo: Juan anduvo cojeando tres kilómetros por el monte desde la Charca de la Olla al Pino Aprisquillo.

Fig. 1: Minimal definition of andar with translated examples



Verbo
Seleccione uno de los verbos de la base de datos

andar ▾

Idiomas
Seleccione entre los idiomas disponibles

☑ Alemán	☑ Árabe	☐ Armenio	☑ Chino
☑ Esloveno	☐ Eslovaco	☑ Finés	☐ Francés
☑ Inglés	☑ Italiano	☑ Japonés	☑ Portugués
☑ Rumano	☑ Ruso		

Verbo: andar

En el verbo 'andar' se han definido 2 acepciones y 1 perífrasis.

Acepción 1	Significado espacial no dinámico: 'hallarse en algún lugar'
Acepción 2	Significado no espacial dinámico: 'hacer algo una y otra vez; experimentar un estado una y otra vez'
Perífrasis 1	andar + GERUNDIO

Gradación: Las acepciones y las perífrasis aparecen ordenadas de menor a mayor según el grado de modificación de la estructura eventiva del verbo.

Pulse en las etiquetas azules

Verbo: andar

Estructura Argumental	[?]
Estructura Temática	[?]
Estructura Eventiva	[?]
Estructura de 'Qualia'	[?]
Modificaciones	[?]

Fig. 2: Lexical entry of *andar* (“Modifications”)

Maria Vittoria Calvi and Luis Javier Santos López

From the Linguaturismo glossary to the Dictionary of Food and Nutrition: proposal for a new electronic multilingual lexicography

Abstract: After addressing certain potentialities of ICT in the field of lexicography, this article will present two works developed at the Department of Language Mediation and Intercultural Communication at the University of Milan. The first one is Linguaturismo, a Spanish-Italian bilingual terminology glossary drawing on a representative corpus of the language of tourism –remarkably, a genre-based approach was adopted in this terminology project. The second one is the Dictionary of Food and Nutrition, with a multilingual terminological basis (Arabic, Chinese, English, French, German, Italian, Portuguese, Spanish and Russian) and specialised in several subthemes proposed for the 2015 Milan Universal Exhibition. The methodology employed in these two products will be emphasised and the main results of both studies discussed, highlighting similarities and differences, with a strong focus on the usefulness of a text genre-based approach in terminology and the positive interrelations between lexicography and terminology.

Keywords: lexicography and terminology, language of tourism, food and nutrition, genre analysis and terminology, electronic multilingual lexicography

1 Limits and potentialities of ICT in lexicography

Nowadays, within the fields of lexicography and terminology, the exponential development of ICT has brought about major changes in project management and, more importantly, in the theoretical paradigm of these two disciplines. Understanding and addressing the limits and potentialities of these changes has become the foremost concern of terminologists and lexicographers.

The impact of today's general economic crisis on the publishing market has been particularly profound. These economic factors have been added to a widespread conception of the Internet as an environment from which all kinds of products can be

Maria Vittoria Calvi: Università degli Studi di Milano, Piazza Indro Montanelli 1, 20099 Sesto San Giovanni (MI), tel. +39 0250321629, fax + 39 0250321640, maria.calvi@unimi.it
Luis Javier Santos López: Università degli Studi di Milano.
This contribution was written in collaboration, before the death of Javier Santos López on 12/12/2016.

https://doi.org/10.1515/9783110607659-006

obtained (leisure, office automation, culture, programming, etc.), either for free or for ridiculously low prices, as evidenced by a significant number of applications for mobile devices.

The fact that some dictionaries by important institutions, such as the Spanish Royal Academy, are freely available to all is not exactly an incentive for private monolingual lexicography projects. Software piracy also discourages publishers from creating electronic products without mid-term data-protection safeguards.

In this context, publishing houses have serious economic doubts when undertaking lexicographic projects. On the other hand, new potentialities are offered by current tools, among which:

a. new funding options derived from publicity;
b. better distribution and segmentation thanks to cyber-shops, which make it possible to reach a universal public;
c. new cost-reducing working methods. Lexicographic editorial offices have disappeared from publishing houses –online work is the norm nowadays, using computer software which significantly reduces staff and infrastructure costs;
d. availability of a wealth of instrumental and documentary resources, greatly simplifying lexicographic management, writing, revision and editing.

On the Internet we can find large corpora in a wide array of languages, freeware corpus analysis toolkits, including AntConc [http://www.laurenceanthony.net/software/antconc/][1], or quite affordable ones, such as WordSmithTools [<http://www.lexically.net/wordsmith/version6/index.html>][2]; dictionary writing toolkits, terminology databanks and glossaries within reach not only of businesses, but also of individuals and researchers, among which Multiterm [<http://www.translationzone.com/products/multiterm-desktop/>][3]; Tlex [<http://tshwanedje.com/tshwanelex/>][4]; or TlTerm [<http://www.ticalc.org/archives/files/fileinfo/41/4143.html>][5]. These resources can be instrumental or documentary. For instance, whereas reviewing a lemma used to change the pagination of a whole dictionary, today's databases allow simultaneous self-editing, writing or revision, and it is possible to paginate just by pressing a key. The typographic complexity of a dictionary, in which every change of letter type, colour or size is significant and not merely for decorative purposes, involved a laborious, slow and costly process when working on paper or with video-writing systems; writers using databases do not make choices regarding typography, as the database offers the default settings.

[1] Last access: March 05, 2017.
[2] Last access: March 05, 2017.
[3] Last access: March 05, 2017.
[4] Last access: March 05, 2017.
[5] Last access: March 05, 2017.

The elements mentioned so far might seem completely unrelated to lexicography. We believe, however, that they have a deep impact on lexicography and, therefore, scholars should factor them in if theoretical discussions are to have a positive influence on professional work. In Sager's words (2002: 30), the products of these disciplines are planned, produced and assessed in terms of a cost-benefit analysis, which is why human, economic and technological aspects should be considered (Maldonado 2012).

Faced with this situation, it is our view that the collaboration of publishers and universities may provide some answers to these limits, while at the same time taking advantage of the potentialities of today's terminology and lexicography. Availability of resources allows us to better adapt to scientific approaches by overcoming the barriers of traditional management. Collaborating with publishing houses, for its part, enables the adequate dissemination of the results of scientific work.

In order to address these issues, from the Department of Language Mediation and Intercultural Communication at Milan University we have engaged in two projects trying to combine methodological and scientific innovation with practical application. These projects cover a long research history which, from the field of terminology applied to tourism in the Spanish-Italian language pair, leads us to the specialised multilingual lexicography of food.

This article aims to review the theoretical basis common to both projects, underline the methodology employed, and highlight the most remarkable features of both products. Having reached the end of the road, we intend to collect and expand the considerations presented in previous work (Bonomi 2014, Bonomi/Santiago González/Santos López 2014, Calvi 2016, Santiago González 2014, Santos López 2011, 2014a, 2014b, 2015, 2016a, 2016b) to emphasise common characteristics and differences, as well as to lay the foundations for future works. As we will now see, the unifying thread that runs through both products has been the study of text genres applied to terminology.

2 The Linguaturismo glossary

The first product proposed is the Linguaturismo glossary, one of the results of the 2007 National Research Project (PRIN) on *El lenguaje de la comunicación turística español-italiano. Aspectos léxicos, pragmáticos e interculturales* ('The language of Spanish-Italian tourism communication. Lexical, pragmatic and intercultural aspects,' 2007ASKNML), funded by the Italian Ministry of Education (MIUR) and coordinated by Maria Vittoria Calvi [http://www.linguaturismo.it/index.htm]. The project was primarily aimed at systematically analysing the genres employed in tourism. Therefore, a hierarchical classification taking account of the variety of (often hybrid) genres was suggested, through which tourism communication is developed, without ignoring the new genres spread by means of the Internet (Calvi 2010).

Our goal was to focus on the most outstanding aspects of the genres created by specialists in tourism, both for specialist-public and public-specialist communication. To that end, we adapted the most usual criteria in genre analysis (see Adam 1999, Bazerman 1994, Berkenkotter/Huckin 1995, Bhatia 1993, 2004 and 2015, García Izquierdo 2007 and 2009, Shiro/Charaudeau/Granato 2012, Swales 1990, among others) to the characteristics of this language, thus providing some guidelines for further developments, as was the case in terminology.

The basic research tool was a bilingual comparable text corpus (Italian and Spanish), including 3,873,664 words, 1,767,905 in Italian and 2,105,759 in Spanish (Mapelli/Piccioni 2011: 47), meeting Biber (1994)'s and Sinclair (1991)'s criteria. A multifunctional, multidimensional approach was adopted in its compilation (García Izquierdo 2007, 2009). As for the genre-based approach, our starting point was the system proposed by García Izquierdo (2007 and 2009) and the GENTT research group for the creation of an encyclopaedia of legal, scientific and economic genres.

According to Calvi's description (2010), the corpus is structured by:

- Genre families, determined by the professional community of origin and its main objectives.
- Macro-genres, defining tangible products which can be identified mostly by the source and channel employed and by the dominant purpose, although characterised by the hybridisation of several genres, text typologies, styles, etc.
- Functionally and formally autonomous genres.
- Subgenres, expressing thematic specification[6].

The Linguaturismo corpus, which served as a basis for our terminological work, is exhaustive and contains text samples of every genre in tourism. As a consequence, our work is not based on an ad hoc corpus but on a sub-corpus drawn from the main one, compiling texts with economic and normative information, programmes, itineraries and tickets, as these, in our expert's opinion, are necessary in tourism management, understood as the work of professionals in wholesale and retail travel agencies. The sub-corpus selected comprises:

- Legal genres: texts focused on tourism regulations, characterised by a high level of specialisation (laws, legislative texts, etc.);

[6] For a more thorough study see Calvi (2010, 2011), where the taxonomy of genres studied is presented in depth.

- Commercial genres (catalogues, programmes, brochures): informative texts aiming to promote sales of tourism products, which are on the semi-specialised communication axis between the tourism industry and users;
- Organisational genres: contracts, tickets, bookings and other internal documents of agencies, comprising the core of communication among specialists in the field and between specialists and clients;
- Institutional and academic genres: reports and papers on tourism. These texts contain economic or sociological information, or discuss tourism from various perspectives (Bonomi 2014: 14).

The innovative method of applying the genre-based approach to terminology should be noted. Even if the literature on genre-based approaches is extremely rich, efforts to apply it to terminology are scarce (Edo Marzá 2012), even non-existent. Text typologies, on the other hand, have been present in terminology methodologies for several years. We opted for the genre-based approach as delving into these genres is essential to effective communication and significantly increases the semantic precision of terms and the control of conceptual structures (Cabré 2002). Furthermore, in contrast to text typologies, genres enable a "grammar of text typologies" (Adam 1999: 34), thus making text selection, and in our case labelling, more exhaustive and consistent. Accordingly, text genre becomes an element not only to organise but also to create knowledge (Adam 1999), as it allows us to go beyond the virtual domain of lexicography or terminology to a real domain of discourse:

> Specifically, the genre label solves problems of total or partial synonymy, as the term is limited to its specialisation level and synonyms are more unlikely to appear. Besides, this label indicates whether an SKU is appropriate in a certain text type depending on the vertical dimension of specialised language, the relation between interlocutors and purpose. It also contributes to understanding and including culturemes in terms of text genre. Finally, genre labels indicate the geographic usage of SKUs.
>
> Translated to English from Santos López 2011: 269

In addition, we should not forget that the latest approaches on the notion of genre (Critical Genre Analysis) include professional practice as a relevant aspect of analysis (Bhatia 2015), thus allowing for a more effective interpretation of the relation between text (and term, in this particular instance) and context.

As a purely terminological product, the theoretical foundations of the Linguaturismo glossary are rooted in the communicative theory of terminology (Cabré 1999) and the resulting "theory of doors", emphasising the need for analysis at the referential, cognitive and linguistic level to conduct a multifaceted analysis of the relations between a concept and its possibilities of terminological treatment (Cabré 2008: 32).

On this basis of this material we proceeded to the strictly terminological work, involving the selection of term candidates, creation of the conceptual map under an expert's supervision, identification of terms and creation of terminology entries comprising the following fields:

- Domain
- Sub-domain
- Term in language A
- Grammatical labelling in language A
- Text genre labelling in language A
- Equivalent in language B
- Grammatical labelling in language B
- Text genre labelling in language B
- Definition
- Contexts in language A
- Contexts in language B
- Synonyms in language A
- Synonyms in language B

The innovative element in this entry is the text genre label, providing pragmatic information on the term which, in our view, translators, writers and professionals in the area should find useful. The label indicates the text genres in which a certain term was found; if the term is present in every text genre, no label appears.

Ex. 1: Entry of *precio del viaje*

precio del viaje (UP) [programa] **costo del viaggio** [normativa] Coste de un desplazamiento por motivos turísticos y de los eventuales servicios incluidos.

Contexto español: El precio del viaje incluye: alojamiento (en hostería, albergue o camping - según detalle de programa), comidas (pensión completa cuatro comidas más refuerzo durante la marcha) salvo los días libres u opcionales, asistencia permanente de guías profesionales de mountain bike. (http://www.palabelonia-biking.com).

Contexto italiano: Più dettagliatamente è solitamente stabilito che la percentuale sul prezzo del viaggio a titolo di costo per il recesso è pari al 10% se avviene circa un mese prima della partenza; al 25% se dalla data dell'effettuazione del viaggio a quella in cui si esercita il diritto intercorrono da 30 a 21 giorni; fine a giungere al trattenimento di circa la metà di quanto versato se avviene da 20 a 10 giorni prima, percentuale che sale al 75% da 10 a 3 giorni e fino alla ritenzione dell'intera somma se si verifica entro i 3 giorni precedenti. (Cistaro, M. 2006. Diritto del turismo e tutele).

Sinónimo italiano: prezzo del viaggio

In example 1 we can see that the term precio del viaje ('travel fare') is used in Spanish in the text genre of programmes, whereas its Italian equivalent 'costo del viaggio' appears in regulations. A translator or writer in Italian who is working on a programme must consider whether the synonym ‘prezzo del viaggio’ should be chosen instead, as in Italian it is also used in programmes. Obviously our glossary is not prescriptive but descriptive, so the genre label does not imply a strict restriction, unlike certain diatopic, diatechnic, diastratic and diaphasic labels in dictionaries – rather, it is a warning for the user to take into account the communicative situations in which a term is preferably used. In traditional glossaries this information can be inferred from contexts, although users find it harder to do, as they must interpret several usage contexts and the query takes significantly longer. In addition, the context selection is not generally representative of the text types employed, and the user will always be doubtful as to whether a certain equivalent may be used in a specific communicative situation.

A label which indicates and systematises the use of a term in a domain and its sub-domains is introduced as a very simple tool to determine the communication levels on the vertical axis of specialised communication, thus completing the horizontal structure of traditional terminology entries by distributing terms according to domains and sub-domains.

In the text genres employed in tourism management little variation was observed, as operators work in an international context, procedures must be standardised for them to be more effective, and infrastructures operate internationally. However, an area of tourism which is highly developed in spite of its recent introduction, rural tourism, allows us to find local and regional realities, for instance casa cueva (‘cave house,’ example 2), casa rural de habitaciones (‘rural house shared with the owners,’ example 3) or casa molino (‘mill house,’ example 4), for which no term exists in Italian; therefore, neological creation was required so that the value of local cultural and economic realities could be highlighted.

Ex. 2: Entry of *casa cueva*

casa cueva (UP) [normativa], [neologismo] **casa in grotta** Alojamiento turístico rural en formaciones rocosas naturales o artificiales.

Contexto español: Esta casa rural se sitúa en una antigua casa-cueva de la ladera del cerro del Castillo de Castrojeriz que se usó como vivienda durante varias décadas.

(http://www.lacasacueva.com).

Creación neológica del terminólogo en italiano.

Ex. 3: Entry of casa rural de habitaciones

casa rural de habitaciones (**UP**) [normativa] **casa rurale condivisa con i proprietary** [neologismo] Alojamiento turístico de arquitectura tradicional, ubicado en zonas rurales y compartido con los propietarios.

Contexto español: En las Casas Rurales de Habitaciones los dormitorios destinados a los clientes deberán estar convenientemente numerados y dotados de cerradura con llave, que deberá entregarse al cliente, o bien cualquier otro sistema que permita el cierre de la habitación y el control del acceso por parte del cliente. (Decreto 243/1999, de 28 de junio, de Navarra).

Creación neológica del terminólogo en italiano.

Ex. 4: Entry of *casa Molino*

casa molino (UP) [normativa] casa mulino [neologismo] Alojamiento turístico situado en un antiguo molino.

Contexto español: El edificio estaba formado por la unión de otras cuatro casas. Se trataba de una Casa-Molino Árabe, en la misma zona se podían encontrar algunas más ya que empleaban el agua de la acequia gorda para mover el Agua de Moler.

(http://www.granadatur.com).

Creación neológica del terminólogo en italiano.

From this standpoint, our glossary refers not only to the industrial dimension of tourism but also to small-scale tourism, always in the context of agency management and therefore excluding relations between individuals. This work also shows the importance of culture when creating specific terminology in areas such as the hotel industry.

The glossary has been published as an annex to the Normas magazine (Bonomi/ Santiago González/Santos López 2014).

3 The *Dictionary of Food and Nutrition* (DFN)

The work devoted to the Linguaturismo glossary served as a starting point for the DFN, especially regarding the genre-based approach applied to lexicography and terminology. The DFN is a specialised dictionary with a terminological basis, multilingual (Arabic, Chinese, English, French, German, Italian, Portuguese, Russian and Spanish) and polyalphabetic (Arabic, Chinese, Latin and Cyrillic), on the subjects of the Milan Universal Exhibition held in Milan in 2015. The DFN resulted from a collective effort, coordinated by Javier Santos, within the research group on specialized lexicography (GRiLS Gruppo di Ricerca sulla Lessicografia Specializzata) active in the

Department. In addition to a scientific committee, an editorial committee was created, comprising a coordinator, a lexicographic and editorial director, a coordinator for each of the languages in the project, and an expert consultant in each of the domains included[7]. The editorial office was mostly formed by students from the Università degli Studi di Milano, in collaboration with students from Milan's Civica Scuola di Traduttori e Interpreti, Cairo's Ain Shams University, the Université Paris-Sorbonne, the Universidad de Valladolid and the Universidad de Murcia.

Even though the DFN contains no publicity, it is the fruit of the collaboration between a public institution, our University, two companies (Autogrill and Ediciones Plan) and a foreign cultural institute (the Confucius Institute).

The title of the Exhibition, reflecting its thematic domain, was "Feeding the Planet, Energy for life"; the seven sub-domains proposed by the organisation stem from it:

– science and technology for food safety and quality;
– science and technology for agriculture and biodiversity;
– innovation in the food chain;
– food education;
– food and lifestyles;
– food and culture;
– cooperation and development in food.

As we can see, the topic list is too long and vague, beyond the scope of a specialised dictionary. Our first task, therefore, was to define the thematic areas which could be terminology domains and create an ad hoc corpus for the subjects included. We limited the scope of the study to four sub-domains on food to which a terminological and lexicographic treatment could be applied. Obviously some of them, such as "food and culture", are too vast and indeterminate to be addressed, and others, such as "food education", introduce a new thematic area, pedagogy, which is not consistent with the others. The domains included are the following:

– gastronomy
– biotechnologies for food
– nutrition
– food safety

[7] Scientific committee: Maria Vittoria Calvi, Giuliana Garzone, Marie-Christine Jullion, Alessandra Lavagnino and Luis Javier Santos López. Lexicographic and editorial director: Loredana Accornero. German language coordinator: Britta Nord. Arabic language coordinator: Francesco De Angelis. Chinese language coordinator: Chiara Bulfoni. French language coordinator: Marie-Christine Jullion. English language coordinators: Chiara Degano and Giuliana Garzone. Russian language coordinator: Paola Cotta Ramusino. Spanish language coordinator: Luis Javier Santos López. Nutrition consultant: Paola De Toni. Biotechnology consultant: Elena Riva. Food safety consultant: Nicola Caramaschi.

Once the domains were determined, and in collaboration with our experts, we created four corpora in each language with the text genres listed in table 1.

Tab. 1: Domains, text genre and sub-domains

Domain: Gastronomy	
Text genre	**Sub-domains**
– Recipe books – Handbooks – Specialised magazines on industrial gastronomy	– ingredients – methods – instruments
Domain: Nutrition	
Text genre	**Sub-domains**
– University handbooks – Norms – Standards – Guidelines – Socio-sanitary dissemination	– Food and nutrients – Modes and measurements for nutrient use – Diet studies – Clinical nutrition
Domain: Biotechnology	
Text genre	**Sub-domains**
– University handbooks – Norms – Information from associations in the area – Essay	– Traditional biotechnology – Genetic engineering – MGS and MGO: processes – applications – Bioethics
Domain: Food safety	
Text genre	**Sub-domains**
– University handbooks – UNI standards – Specialised magazines – Dissemination	– Risks – biological – physical – chemical – product composition, packaging and labelling

The text genres within each domain were chosen because they are the most representative ones of the industry, and also because the potential user of our dictionary is undetermined, unlike that of the Linguaturismo glossary. Our dictionary is aimed at visitors of the exhibition, approximately twenty million people, according to the organisers. We do not know how well versed potential users are in these topics or how they will use the dictionary –we assume they will use it mostly for decoding, but coding needs cannot be ruled out. The texts selected as being representative of each domain belong to widely or reasonably disseminated genres, excluding those targeted at specialists, such as scientific papers.

Besides these methodological premises, in the case of the DFN our work was based on the socio-cognitive approach (Temmerman 2000) as it allowed us to assign notions to the concept network and enabled term relations as a cross-sectional structured event across different categories. This is a key aspect in this work, as we are working with four different domains and many terms are cross-sectional and belong to several of them.

Given that we have created a lexicographic product with terminology databases, much attention was paid to the relationship between lexicography and terminology, a problem solved in theory but not yet in practice. As already stated our potential users are not specialists, so we followed a translation-based trend, without taking account of planning and standardising trends (Cabré 1993: 45). At a practical level we worked with a middle-out approach, combining the study of terms in their communication contexts (bottom-up approach) with an initial categorisation developed by experts (Durán Muñoz 2012: 196, Santos López 2015).

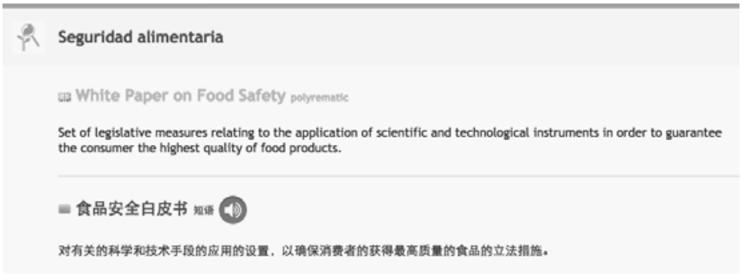
We will now see the main problems faced when conducting the project and how they were solved.

3.1 No representative texts of some domains (biotechnologies for food and food safety) in some languages (Arabic, Russian, Chinese)

Compiling this multilingual corpus did not pose many problems in the domains of gastronomy and nutrition, but it did in the other two. The domain of biotechnology for food gave rise to serious problems in Arabic because it was really difficult to find representative texts in this language, as English is the dominant vehicle for communication in this area. Food safety is a highly developed discipline in the framework of the European Union, which includes many of the languages in our dictionary, but not in other linguistic areas, such as the Chinese, Russian or Arabic ones. We were unable to create a comparable corpus in these languages. We had no problems to find a text corpus on physical, chemical or biological risks, but it was hard to find one, for instance, on safety procedures, such as the well-known HACCP[8] (Hazard Analysis Critical Control Point), documentation, for instance "libro blanco sobre seguridad alimentaria" ('white paper on food safety,' example 5), steps in processes, such as "punto crítico de control" ('critical control point,' example 6), procedures, like "variación de aceptabilidad" ('acceptability variation'), etc.

[8] International protocol on food contamination prevention throughout the production process.

Ex. 5: Entry of White Paper on Food Safety



Seguridad alimentaria

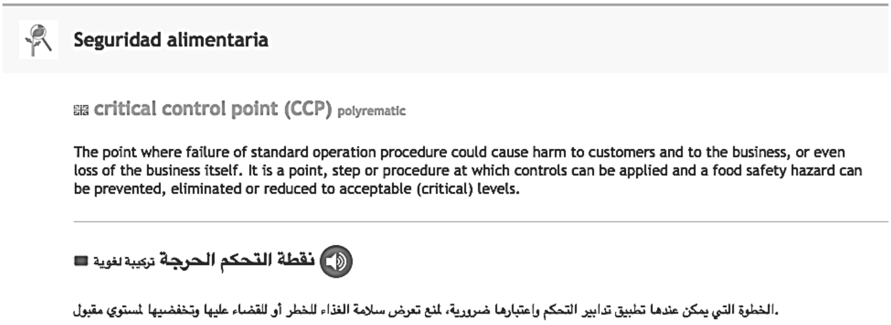
White Paper on Food Safety polyrematic

Set of legislative measures relating to the application of scientific and technological instruments in order to guarantee the consumer the highest quality of food products.

食品安全白皮书 短语

对有关的科学和技术手段的应用的设置，以确保消费者的获得最高质量的食品的立法措施。

Ex. 6: Entry of critical control point (CCP)



Seguridad alimentaria

critical control point (CCP) polyrematic

The point where failure of standard operation procedure could cause harm to customers and to the business, or even loss of the business itself. It is a point, step or procedure at which controls can be applied and a food safety hazard can be prevented, eliminated or reduced to acceptable (critical) levels.

نقطة التحكم الحرجة تركيبة لغوية

الخطوة التي يمكن عندها تطبيق تدابير التحكم واعتبارها ضرورية، لمنع تعرض سلامة الغذاء للخطر أو للقضاء عليها وتخفضيها لمستوي مقبول.

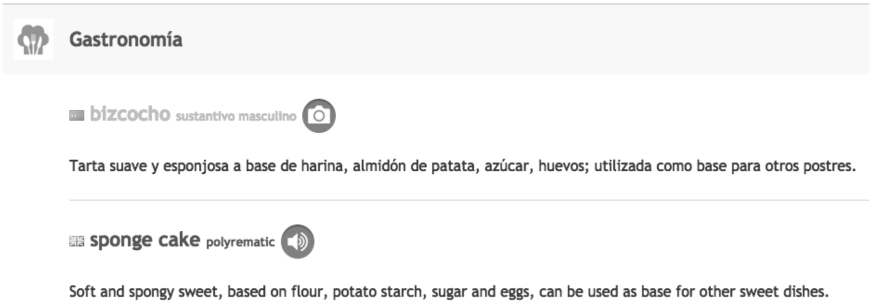
At this first stage of the creation of the dictionary, with a terminology-based approach, we proceeded to extract terms, select term candidates, and contrast the languages studied. The most technical domains posed no problems, although we could not complete our terminological work on food safety due to the aforementioned lack of corpora in some languages. This is why, also in this domain, we employed lexicographic tools to complete our work.

3.2 Difficulty in demarcating the domain of gastronomy (no dishes, only ingredients, methods and procedures were taken into account)

The domain of gastronomy has proved to be impossible to address from a strictly terminological approach for two reasons.

In the first place, the broad scope of the domain, covering the whole food chain up to the table, forced us to limit this field to the kitchen. In fact, terms such as mantel (‘tablecloth’) or servilleta (‘napkin’), used at the table, or carnicería (‘butcher’s’), which –being part of the food chain as a distribution point, albeit previous to the kitchen– belongs to gastronomy, are not in the dictionary. We also had to restrict the domain to ingredients, methods and procedures, since, for example, including the names of elaborate dishes would have involved a string of terms which do not exist in the other languages. This is the case of the term fabada ('Asturian stew'), with equivalents in the remaining languages which are merely a calque or a description containing the essential elements of the definition. A dictionary in nine languages is so complex that this type of solution would be used excessively, which would result in too much information (Tarp 2007) and would be detrimental to its usability. Only names of elaborate products which are ingredients of other dishes as well were included, such as bizcocho ('sponge cake'), an ingredient employed to bake cakes (example 7).

Ex. 7: Entry of *bizcocho*



Gastronomía

bizcocho sustantivo masculino

Tarta suave y esponjosa a base de harina, almidón de patata, azúcar, huevos; utilizada como base para otros postres.

sponge cake polyrematic

Soft and spongy sweet, based on flour, potato starch, sugar and eggs, can be used as base for other sweet dishes.

3.3 Diatopic variation in gastronomy

The second problem we were faced with was diatopic variation in gastronomy. As an example, without considering the Latin American variation, what is usually called atún blanco ('white tuna') has other commercial names in Spain, such as bonito del norte and albacora, in Asturias it is commercialised as mono, in the Balearic Islands as ullada, in the Canary Islands as barrilote and in the Basque Country as hegalabur. If this variation had been multiplied by the nine languages in the dictionary, our work would have been endless.

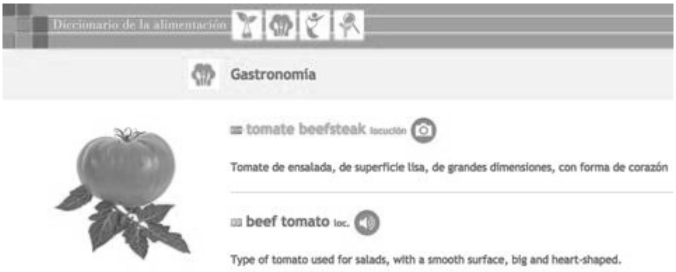
In summary, the problem posed by gastronomy is that it is a specialised language, as a professional language, which almost completely coincides with common language. Common language cannot be treated with a terminology-based but with a lexicography-based approach, and consequently we had to change it and adopt what we called "lexicoterminology," that is, lexicography with a terminological foundation. This turned out to be really useful for writing the dictionary and checking its consistency. Thus, we did not shy away from creating conceptual maps of the different cultures, which were really helpful when establishing the macrostructure of the dictionary and writing the definitions. Likewise, the extraction of defining and usage contexts was of high importance in writing definitions and validating equivalences.

The final format of this work is lexicographic as well, as a bilingual specialised dictionary, since each entry comprises the following fields:

- domain
- lemma in language A
- definition in language A
- picture
- part of speech in language A
- lemma in language B
- part of speech in language B
- definition in language B
- pronunciation in language B (audio)

This format is peculiar in that it is multimedia, as it includes pictures and audio files with the pronunciation of lemmas in the nine languages. In example 8 the Spanish-Chinese article of tomate beefsteak (‘beefsteak tomato’) is shown, in which, if the picture is activated, the result is more clarifying than the definition of the product (‘Salad tomato, with a smooth surface, large in size, heart-shaped’). It has not been possible to document every lemma with pictures because some concepts are abstract, immaterial, etc, and cannot be illustrated. Other concepts, such as ADN bicatenario (‘double-stranded DNA’), are material, but a picture would give absolutely no information to the potential user of the dictionary.

Ex. 8: Entry of *tomate beefsteak*



Gastronomía

tomate beefsteak locución

Tomate de ensalada, de superficie lisa, de grandes dimensiones, con forma de corazón

beef tomato loc.

Type of tomato used for salads, with a smooth surface, big and heart-shaped.

4 Discussion of results

Above all, the major differences between the two products elaborated by us should be highlighted, as shown in Table 2, to proceed to a comparative analysis of the results.

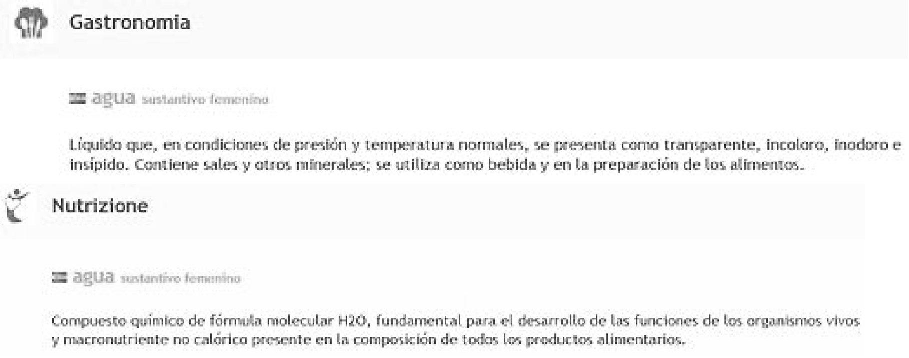
Tab. 2: Comparative analysis of the results

Linguaturismo Glossary	**DFN**
Based on sub-corpora drawn from a general corpus	Based on an ad hoc corpus
Genre selection depending on the target reader	Genre selection depending on the indeterminacy of the target reader
Study of genres for the general purposes of the project	Study of genres applied to terminological work
Management of variation	Impossibility of managing variation
Language policy: valuing minority cultures	Language policy: valuing minoritised languages

Selecting a corpus drawn from the general corpus used in Linguaturismo allows us to specifically situate the potential user within a broader communicative setting. Term candidates are included or excluded by studying the sub-corpus, and also using the main corpus as reference corpus. By observing how frequent and widespread term candidates are, we can decide on the specificity and need for inclusion of a certain candidate for the professional purposes established beforehand. In the case of the DFN, on the other hand, its complexity, scope of the domains and indeterminacy forced us to create a non-exhaustive corpus of the specialised domain, which is merely representative of the most common usage in different thematic areas.

The study of genres we undertook in Linguaturismo is indebted, as already mentioned, to the study of the whole corpus, which allowed us to establish a general taxonomy of the language of tourism. A consequence of this is that we could employ text genre labels at the output stage, and not just at the input stage, as rigorously as is required in scientific work. For the DFN, we studied the genres previously selected on our experts' recommendation, and therefore exhaustiveness is but a hypothesis and representativeness is guaranteed by extra-linguistic criteria. Thus, we could not apply the genre label directly to the lexicographic article, as we were not absolutely certain as to whether we would delete some relevant ones. The study of genres proved to be useful when writing definitions, adapted not only to the thematic area but also to users' estimated competence. Below are, for instance, the definitions of agua potable ('drinking water') in two different sub-domains (example 9).

Ex. 9: Entry of *agua*



Gastronomia

agua sustantivo femenino

Líquido que, en condiciones de presión y temperatura normales, se presenta como transparente, incoloro, inodoro e insípido. Contiene sales y otros minerales; se utiliza como bebida y en la preparación de los alimentos.

Nutrizione

agua sustantivo femenino

Compuesto químico de fórmula molecular H2O, fundamental para el desarrollo de las funciones de los organismos vivos y macronutriente no calórico presente en la composición de todos los productos alimentarios.

In the first case, gastronomy, our user has no specific competence, so the term is described in plain language and its use in the reference field is specified. As for the definition of water in nutrition, we include the indication “for the development of living organisms”, and also its molecular form as a chemical compound. The definition is still transparent, albeit more complex and terminologically dense in the second instance (‘living organism’, ‘macronutrient’, ‘non-caloric’), since we understand that nutrition, as a scientific or professional discipline, requires users to have a greater knowledge of terminology. This is why a more complex definition provides information here, whereas in gastronomy it would cause incomprehension.

The issue of linguistic variation was addressed in Linguaturismo as it was restricted to highly precise areas, such as the aforementioned rural tourism or regional institutions. We can thus see that in Italy rural tourism businesses are classified using spighe (‘sprigs’), whereas in Spain hojas (‘leaves’) would be employed. The term in Italian, however, is valid only for the region of Tuscany, but it can be included in our glossary because of its usage restrictions (example 10).

Ex. 10: Entry of *hoja*

hoja (s. f.) [normativa] **spiga** [programa] Símbolo que indica el nivel cualitativo de la vivienda rural.

Contexto español: No obstante con independencia de la puntuación obtenida, la presencia de alguna de las circunstancias que conllevan la clasificación de 1 Hoja, implicará la asignación de dicha categoría. Igualmente, la presencia de una circunstancia que impide la obtención de la categoría de 3 Hojas conllevará la clasificación del establecimiento como de 1 Hoja o 2 Hojas, de acuerdo con la puntuación obtenida. (Decreto 243/1999, de 28 de junio, de Navarra).

Contexto italiano: Forse vi sarà capitato di vedere, sui siti web di molti agriturismi, un simboletto composto da una, due, tre, quattro o cinque spighe. Non si tratta di un logo o di un elemento decorativo, ma di una classificazione analoga alle stelle degli hotel. (http://www.lecolombe.com).

Denominación italiana usada en la región de Toscana.

In Spain, autonomous communities are usually in charge of tourism management legislation, so we find concepts such as “zona de interés turístico preferente” (‘preferential tourist interest area’), only present in the regulations of some regions (Castile and León in this case). Diastratic variation is specified in the text genre label. In the DFN diastratic, diatopic and even diaphasic variation is huge in the field of gastronomy. Therefore, in our view, it was beyond the scope of a dictionary with combinations in nine languages.

When describing the materials we pointed out that the Linguaturismo glossary included notions not completely established in specialised language, but quite representative of emerging cultures. As a consequence, we endeavoured to add terminology from two minority industries, rural tourism and sustainable tourism, in such a way that it could explain not only the social reality of a time, but also an approach to potential developments in the future. In the DFN we were unable to take account of cultural diversity due to the dimensions and depth of the cultures treated. Nevertheless, we did want to highlight the value of minoritised languages, such as Arabic, or of languages in which some tools could not be accessed because of their different technological development (Chinese, Russian and Arabic in the domain of food safety).

5 Conclusion

In the introduction we talked about overcoming the limits and taking advantage of today’s opportunities in terminology and lexicography. The first issue discussed was economic. After our experience we conclude that, in a university setting, collaboration between institutions and private companies is essential. The Linguaturismo glossary will be disseminated to the extent that it is included in more commercially widespread documents; in fact we are in talks with members of the Cometval project to integrate it into their multilingual products. This project, conducted at the University of Valencia and coordinated by Julia Sanmartín, created the Multilingual Dictionary of Tourism (http://tourismdictio.uv.es/glosario.php) in Spanish, English, French and Arabic, based on the vocabulary and discourse analysis of parallel and comparable corpora of websites with tourism information. Our collaboration may add a fourth language, Italian, to the dictionary.

As for the DFN, the Confucius Institute and Autogrill Spa provided funding and scientific collaboration, whereas the publisher chose to commercialise it as an app for mobile devices and computers so that our product could be as widespread as possible. Business participation in institutional projects requires engagement, sometimes renunciation and specific times, and institutions must ensure their adaptation to the needs of this collaboration. Private funding is possible but appealing products are a requisite in order to spark companies’ interest in participating.

From a scientific and methodological standpoint we have seen that the possibilities of corpus linguistics, the text genre-based approach and a heterodox view of the relationship between terminology and lexicography, allow us to create innovative and effective terminological and lexicographic products fulfilling two key requirements of these disciplines nowadays: being user-oriented and having capacity of collaboration and feedback.

With regard to future developments, it is our belief that the findings discussed herein may encourage promising methodological approaches based on the notion of text genre, opening terminological work to the depth of discourse: whereas the concept of "domain" is restricted to themes, that of "genre" covers discursive practices of specialists in a certain field.

Furthermore, the contribution to ICT-aided multilingual lexicography should be noted. In a globalised world fluid knowledge exchange between languages becomes ever more necessary. A model such as the one described in this paper, combining scientific rigour with accessibility, allows for innovative lexicographic resources which, among other social functions, enhance dissemination of specialised knowledge in a wide array of languages, thus counterbalancing the overwhelmingly dominant position of English.

6 Bibliography

Adam, Jean-Michel (1999): Linguistique textuelle – Des genres de discours aux textes. Paris: Nathan.

Bazerman, Charles (1994): Systems of genres and the enactment of social intentions. In: Freedman, Aviva/Medway, Peter (eds.): Genre and the New Rhetoric. London: Taylor & Francis, 79–101.

Berkenkotter, Carol/Huckin, Thomas N. (1995): Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power. Hillsdale, NJ: Lawrence Erlbaum.

Bhatia, Vijay K. (1993): Analysing Genre: Language in Professional Settings. London: Longman.

Bhatia, Vijay K. (2004): Worlds of Written Discourse. London: Continuum.

Bhatia, Vijay K. (2015): Critical Genre Analysis: Theoretical Preliminaries. In: Hermes – Journal of Language and Communication in Business 54, 9–20.

Biber, Douglas (1994): Using register-diversified corpora for general language studies. In: Armstrong, Susan (ed.): Using language corpora. Cambridge: The MIT Press, 180–201.

Bonomi, Milin/Santiago González, Paula/Santos López, Luis Javier (eds.) (2014): Glosario español-italiano sobre la gestión del turismo. Valencia, Anejo n. 6 de la revista Normas (https://www.uv.es/normas/2014/anejos/Libro_glosario_2014.pdf).

Bonomi, Milin (2014): Linguaturismo, un glosario de gestión del turismo italiano-español. In: Bonomi, Milin/Santiago González, Paula/Santos López, Javier (eds): Glosario español-italiano sobre la gestión del turismo, Anejo n. 6 de la revista Normas, 2–27.

Cabré, María Teresa (1993): La terminología. Teoría, metodología, aplicaciones. Barcelona: Antártida/Empuries.

Cabré, María Teresa (1999): La terminología. Representación y comunicación: elementos para una teoría de base comunicativa y otros artículos. Barcelona: IULA. Universitat Pompeu i Fabra.

Cabré, María Teresa (2002): Sur la représentation mentale des concepts: bases pour une tentative de modélisation. In: Bejoint, Henry/Thoiron, Philippe (eds.): Le sens en terminologie. Lyon: Presses Universitaires de Lyon, 20–30.

Cabré, María Teresa (2008): Realidad, cognición y lenguaje: la poliedricidad como principio. In: Atti del convegno nazionale Ass.I.Term: Terminologia, analisi testuale e documentazione nella città digitale. AIDA informazioni, 1-2, 11–24.

Calvi, Maria Vittoria (2010): Los géneros discursivos en la lengua del turismo: una propuesta de clasificación. In: Ibérica 19, 9–32.

Calvi, Maria Vittoria (2011): Pautas de análisis para los géneros del turismo. In: Calvi, Maria Vittoria/Mapelli, Giovanna (eds.): La lingua del turismo. Bern: Peter Lang, 19–46.

Calvi, Maria Vittoria (2016): Léxico de especialidad y lengua del turismo. In: Duffé Montalván, Aura Luz (ed.): Estudios sobre el léxico: puntos y contrapuntos. Berna: Peter Lang, 187–214.

Durán Muñoz, Isabel (2012): La ontoterminología aplicada a la traducción. Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores. Bern: Peter Lang.

Edo Marzá, Nuria (2012): Lexicografía especializada y lenguajes de especialidad: fundamentos teóricos y metodológicos para la elaboración de diccionarios especializados. In: Lingüística 27, 98–114.

García Izquierdo, I. (2007): Los géneros y las lenguas de especialidad. In: Alcaraz Varó, Enrique/Mateo Martínez, José Mateo/Yus Ramos, Francisco (eds.): Las lenguas profesionales y académicas. Barcelona: Ariel, 119–125.

García Izquierdo, I. (2009): Divulgación médica y traducción: el género información para pacientes. Bern: Peter Lang.

Maldonado, Concepción (2012): Los diccionarios en el mundo ELE: ayer, hoy y mañana (una reflexión desde la propia experiencia). In: Revista Internacional de Lenguas Extranjeras 1, 151–179.

Mapelli, Giovanna/Piccioni, Sara (2011): Taxonomía de los textos turísticos: factores lingüísticos y factores contextuales". In: Calvi, Maria Vittoria/Mapelli, Giovanna (eds.): La lengua del turismo. Bern: Peter Lang, 47–74.

Sager, Juan Carlos (2002): La terminología y la traducción en la sociedad de la información. In: Alcina Caudet, María Amparo/Gamero Pérez, Silvia (eds.): La traducción científico-técnica y la terminología en la sociedad de la información. Castellò de la Plana: Publicacions de la Universitat Jaume I, 17–44.

Santiago González, Paula de (2014): De la forma al contenido, del contenido a la definición: el glosario linguaturismo. In: Bonomi, Milin/Santiago González, Paula/Santos López, Luis Javier (eds.): Glosario español-italiano sobre la gestión del turismo. Valencia, Anejo n. 6 de la revista Normas: 28–44.

Santos López, Luis Javier (2011): El glosario Linguaturismo: aplicación del enfoque de géneros a la terminología. In: Calvi, Maria Vittoria/Mapelli, Giovanna (eds.): La lingua del turismo. Bern: Peter Lang, 249–272.

Santos López, Luis Javier (2014a): El género textual en la terminología turística. In: Bonomi, Milin/Santiago González, Paula/Santos López, Luis Javier (eds.): Glosario español-italiano sobre la gestión del turismo. Valencia: Anejo n. 6 de la revista Normas, 45–69.

Santos López, Luis Javier (2014b): La terminologia dell'Expo 2015: approcci metodologici. In: Jullion, Marie-Christine/Cattani, Paola (eds.): Les langues, les cultures, la traduction pour la médiation: perspectives d'enseignement et de recherche. Paris: L'Harmattan, 81–104.

Santos López, Luis Javier (ed.) (2015): Dizionario dell'alimentazione. Loreto AN: Edizioni Plan/Academia Universa Press Srl.

Santos López, Luis Javier (2016a): Análisis de las relaciones entre terminología, lexicografía y tecnología: el proyecto "Parole per mangiare". In: Cuadernos AISPI 6, 95–110.

Santos López, Luis Javier (2016b): El diccionario de la alimentación: un proyecto didáctico para la enseñanza de segundas lenguas y competencias lexicográficas. In: Chine, Dalila/Pujol Berché, Mercè/Taillot, Allison (eds.): La formación de profesores de español en contextos profesionales, Cahier du GERES 8, 123–137.

Shiro, Martha/Charaudeau, Patrick/Granato Luisa (eds.) (2012): Los géneros discursivos desde múltiples perspectivas: teorías y análisis. Madrid/Frankfurt am Main: Iberoamericana/ Vervuert.

Sinclair, John (1991): Corpus, concordance, collocation. Oxford: Oxford University Press.

Swales, John (1990): Genre Analysis. Cambridge: Cambridge University Press.

Tarp, Sven (2007): ¿Qué requisitos debe cumplir un diccionario de traducción del siglo 21? In: Fuertes Olivera, Pedro Antonio (ed.): Problemas lingüísticos en la traducción especializada. Valladolid: Universidad de Valladolid, 227–256.

Temmerman Rita (2000): Towards New Ways of Terminology Description. The Sociocognitive-approach. Amsterdam/Philadelphia: Benjamin‘s.

Gloria Corpas Pastor and Isabel Durán-Muñoz

INTELITERM: In search of efficient terminology lookup tools for translators

Abstract: There is currently a pressing need to develop specific applications for translators as final users, with the purpose of fulfilling their particular professional requirements. Corpora and advanced lookup options bring benefits to translation and open up a wealth of opportunities in research. This paper presents Inteliterm, an innovative lexicographic resource which combines corpus management tools with different types of searches and customisation options in order to enhance translation results and minimise translators' efforts when searching for terminology. Section 1 provides a brief foreground glimpse of the project rationale. Section 2 delves into the comprehension assistants developed in the 90's as a first step towards present-day intelligent dictionaries. Section 3 provides the theoretical foundation of the novel tool Inteliterm. Beside its proactive translation support functions, this web application also provides a TBX (TermBase eXchange) termbase editor, which allows users to create, edit or upload terminological databases in the standard-format TBX (ISO 30042: 2008) and to query their own databases using the Inteliterm tool. Users' responses and assessment of the tool are also provided in Section 4. Finally, Section 5 includes some concluding remarks with suggestions for further improvements.

Keywords: terminology lookup, translation, corpus management, term management, e-resource, next-generation intelligent dictionary

Gloria Corpas Pastor: Universidad de Málaga/University of Wolverhampton, Facultad de Filosofía y Letras, Dpto. Traducción e Interpretación, Avda. Louis Pasteur, 29071 Málaga (Spain), tel. +34 952133409, gcorpas@uma.es
Isabel Durán Muñoz: Universidad de Córdoba, Facultad de Filosofía y Letras, Dpto. Filología Inglesa y Alemana, Plaza del Cardenal Salazar, 3, 14071 Córdoba (Spain), tel. +34 957218426, iduran@uco.es

https://doi.org/10.1515/9783110607659-007

1 Introduction

There is a pressing need to develop new tools that meet translators' needs satisfactorily (Bowker/Corpas Pastor 2015). Terminology resources face serious limitations as regards their typology, coverage, functionalities and degree of translators' satisfaction (cf. Durán-Muñoz 2010). Against this background, this section presents *Inteliterm*[1], an integrated and advanced corpus-management and lookup tool. This novel software application has been developed as a suitable and productive tool for professional translators, bearing in mind their different needs and expectations.

Inteliterm follows De Schryver's proposal (2009) of building flexible and intelligent dictionaries that would be able to "study and understand its users". This new type of 'dictionary' is deeply rooted in the principles of LSP lexicography, whose main purpose is to develop lexicographical tools that function as utility products that provide specific types of help to specific types of user in specific types of user situation related to one or more subject fields and their LSP (Nielsen 2018: 71). Thus, it is a dynamic and functional electronic dictionary aimed at assisting translators when understanding the source text and producing the target text in innovative ways, mainly concerning customisation and access possibilities through user profiling (cf. Bergenholtz 2011, Spohr 2011, Tarp 2011, Nielsen 2018). The result is a multilingual terminology resource in the domain of health and beauty tourism and is aimed at translators as potential end users.

2 The core: intelligent dictionaries

The first applications that could be considered "intelligent dictionaries" are the so-called *comprehension assistants* (assistants for decoding/understanding a foreign language), which emerged in the 1990s. These tools were intended to facilitate the understanding and decoding of a message written in another language by users with little to no knowledge of that language, in addition to streamlining the lookup of the meaning of words in the text without the need to consult other external sources. At the beginning, they appeared as an alternative to machine translation so as to remedy the frequent shortcomings and errors committed by this type of trans-

[1] This application has been designed and developed within the framework of the Spanish RD project INTELITERM: Sistema inteligente de gestión terminológica para traductores [INTELITERM: Intelligent Terminology Management System for Translators] (ref. no. FFI2012-38881, 2012-2016). It is a multilingual tool (Spanish, English, German and Italian) for the domain of health and beauty tourism, but could be extended to other language combinations and specialised domains. It has also been partially supported by the Spanish Ministry of Economy and Competitiveness (grant no. FFI2016-75831P) and the Andalusian Regional Government (grant no. HUM2754).

lation. They also presented themselves as more flexible and agile tools than conventional general dictionaries.

Unlike dictionaries and other resources, either electronic or paper-based, which require the insertion or selection of words in the resource, these applications allow the looking up of information (equivalents, mainly) directly by clicking on the desired words in the working text. These tools have been developed within computational linguistics and natural language processing projects and, thus, they have frequently employed pre-made sources of information as resources, such as WordNet[2], and do not have specific user groups in mind but general public as well as general language. By way of example, we can mention some tools of this kind, such as the prototype COMPASS (Feldweg/Breidt 1996), which was the first one in the series of comprehension assistants to have been developed in recent decades. Newer and more complete versions of these assistants have been called "intelligent dictionaries", such as *Sharp Intelligent Dictionary* (SID) (Whitelock/Edmonds 2000), *Intelligent Dictionary Help System* (IDHS) (Agirre et al. 2003), *MobiMouse Plus* (Prószéky/Földes 2005: 2), *Smarty* (Arnaudov/Mitkov 2008) and Dixio[3].

Although intelligent dictionaries represent a step forward compared to the rest of the resources that pursue a similar end, the fact is that they have relatively limited value due to their coverage of language and users, that is –as already mentioned – general language and unskilled users, respectively. In fact, the intelligent dictionaries mentioned above only present more or less sophisticated functions for textual disambiguation, usually for a limited number of language pairs and general language. To the best of our knowledge, even the most advanced ones, such as Smarty, are not "suitable" for other register-specific uses and, therefore, are far from fulfilling translators and interpreters' needs working in a specialised domain. They also do not include other features that could alleviate their lack of specialisation, such as the insertion of new databases, (conventional or ad hoc) glossaries, external searches or corpus management tools. These and other limitations make current intelligent dictionaries unsuitable for specialised communication and translation.

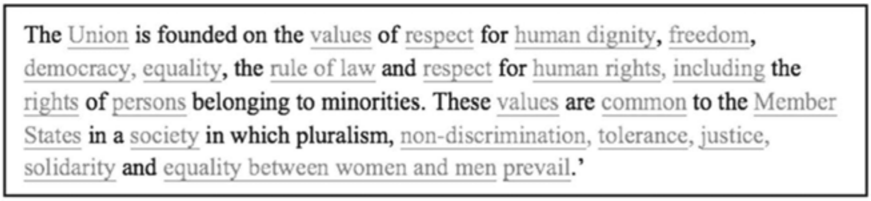
Some other attempts have also been carried out so far: *Terminator*, within the Evroterm project[4] (Željko 2009), and Trandix (Durán-Muñoz 2014, Durán-Muñoz/Fernández Sola 2014). These applications represent a step forward; they have been developed as a more sophisticated application aimed at translators and based on a

[2] [<https://wordnet.princeton.edu/>; last access: October 20, 2018].

[3] This application is commercial and can be acquired at the online shop: [<http://shop.dixio.com/en/>; last access: May 15, 2017], where there is also a demo version available.

[4] The Evroterm project's purpose is to develop a multilingual terminology database in 15 languages (although most of the equivalents are only in English and Slovenian) that would support translators from the Slovenian Unit of Translation Government Office for European Affairs (GOEA) of the Republic of Slovenia during the process of translating the European Union documentation, prior to joining the institution.

specialised domain. Terminator is a terminological analyser (or what has been considered until now a comprehension assistant or intelligent dictionary) where the user can enter a text (written in English, German, French or Slovenian), and the application, once the text is processed, marks with hyperlinks the terms found in the Evroterm database (EDB). This application has a very simple and intuitive interface, making it easy to use by potential users.



The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail.'

Fig. 1: Terminological analysis by *Terminator*

As can be seen in Figure 1, many terms from the working text are marked, allowing users to quickly and easily consult the terminological information about those terms included in the EDB. Given the context in which the tool was developed (the translation of the minutes of the European Union for the inclusion of Slovenia in the European institution), the predominant terminology relates to legislation developed within the European Union and includes mainly equivalents for the Slovenian-English combination. However, this tool has some drawbacks that hinder efficient and effective use by translators, such as the limited amount of information included in the database (only context, equivalent and, in some cases, definitions) and the lack of flexible and updated information.

Trandix is a multilingual terminological resource (Spanish, English and German) focused on the domain of adventure tourism and with translators as prospective end users. As with *Terminator*, *Trandix* facilitates the search in context by highlighting the terms included in the internal database of a working text, but it also permits the customisation of terminological entries and other features, and allows users to launch searches from the application to external resources in order to extend the information found in the tool or to find information not included in the resource. Moreover, the tool also provides the opportunity to send feedback by users to the developers with the aim of communicating problems, lack of information, inconsistencies, etc. In short, the diverse search possibilities provided by *Trandix*, along with its other features, help speed up translators' research tasks by saving their time and effort, especially when forced to use and consult many different resources to find relevant information.

Despite the advantages and improvements brought by these tools, there is still the need for a comprehensive application that overcomes their shortcomings and increases their usability in translation. As discussed before, the incorporation of corpus management and other NLP tools, such as term extractors or aligners, would be of great use to ease and speed access and lookup and to provide useful pragmatic information for translators in the form of concordances.

Taking into account the needs of our end-user group, and based on the previous and successful systems –especially the *Trandix* tool, which was previously developed by the authors of this paper, we designed and developed a novel application, *Inteliterm*, which is described in more detail in the following sections.

3 A step forward: main functionalities

As previously stated, the purpose of this research was to develop a flexible and integrated tool which addresses users' needs in specialised communication, and more specifically professional translators (a very important group of target users within the framework of LSP lexicography). Our main objective was the design and implementation of a system whose architecture is similar to an intelligent multilingual dictionary but, at the same time, has a modular and open structure that enhances the performance of its users by means of better access, lookup and particularly results in translation tasks.

This intelligent modular system combines an efficient lookup system through several types of searches, terminology as well as corpus management tools, with the aim of addressing translators' needs and reducing the fruitless time devoted to terminology research. Another important aim of this tool is to promote constant updating and improvement based on direct feedback by users.

The application is composed of three different modules, all of which offer different functionalities and customisation according to user preferences: 1) Term search; 2) Corpus management tool; and 3) Terminology management tool.

Despite its modularity, the application provides a user-friendly and intuitive interface, which makes it very easy to use. Moreover, as it is web-based, users can access *Inteliterm* from anywhere and at any time. It also permits multiple users and operational systems, including a user manual that provides information regarding all its functionalities.

The working process is very simple: basically, the user loads the working text, either in a file or copy-pasting it, and processes the text (Figure 2). During this process, which lasts only a few seconds, the uploaded text is lemmatised with TreeTagger, and the terms that are found in the health and beauty terminological database are marked with hyperlinks on the text[5] (Figure 3). The tool automatically detects the language of the text so the marking process is more effective. At this point, the user can edit the text in the same application, copy-paste and undo as if he/she were in a text editor, and process the text as many times as needed.

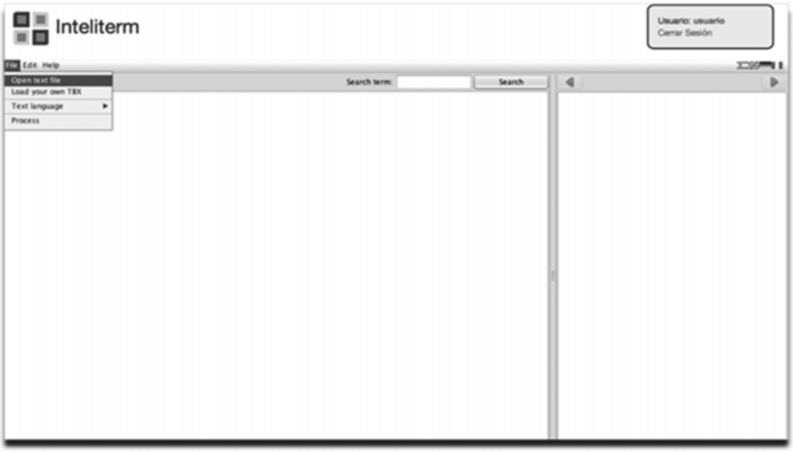
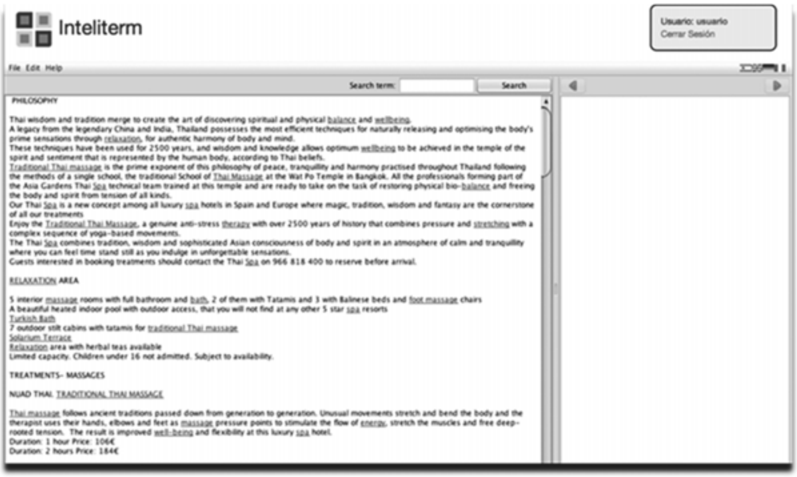


Fig. 2: Loading a text file



PHILOSOPHY

Thai wisdom and tradition merge to create the art of discovering spiritual and physical balance and wellbeing.
A legacy from the legendary China and India, Thailand possesses the most efficient techniques for naturally releasing and optimising the body's prime sensations through relaxation, for authentic harmony of body and mind.
These techniques have been used for 2500 years, and wisdom and knowledge allows optimum wellbeing to be achieved in the temple of the spirit and sentiment that is represented by the human body, according to Thai beliefs.
Traditional Thai massage is the prime exponent of this philosophy of peace, tranquillity and harmony practised throughout Thailand following the methods of a single school, the traditional School of Thai Massage at the Wat Po Temple in Bangkok. All the professionals forming part of the Asia Gardens Thai Spa technical team trained at this temple and are ready to take on the task of restoring physical bio-balance and freeing the body and spirit from tension of all kinds.
Our Thai Spa is a new concept among all luxury spa hotels in Spain and Europe where magic, tradition, wisdom and fantasy are the cornerstone of all our treatments
Enjoy the Traditional Thai Massage, a genuine anti-stress therapy with over 2500 years of history that combines pressure and stretching with a complex sequence of yoga-based movements.
The Thai Spa combines tradition, wisdom and sophisticated Asian consciousness of body and spirit in an atmosphere of calm and tranquillity where you can feel time stand still as you indulge in unforgettable sensations.
Guests interested in booking treatments should contact the Thai Spa on 966 818 400 to reserve before arrival.

RELAXATION AREA

5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath
7 outdoor stilt cabins with tatamis for traditional Thai massage
Solarium Terrace
Relaxation area with herbal teas available
Limited capacity. Children under 16 not admitted. Subject to availability.

TREATMENTS- MASSAGES

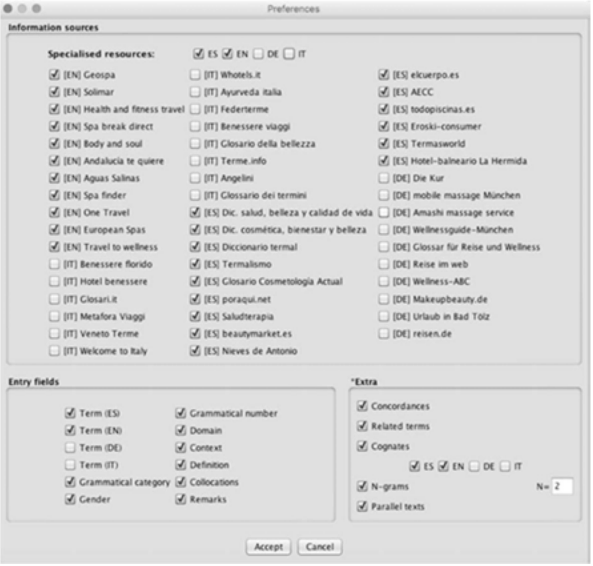
NUAD THAI. TRADITIONAL THAI MASSAGE

Thai massage follows ancient traditions passed down from generation to generation. Unusual movements stretch and bend the body and the therapist uses their hands, elbows and feet as massage pressure points to stimulate the flow of energy, stretch the muscles and free deep-rooted tension. The result is improved well-being and flexibility at this luxury spa hotel.
Duration: 1 hour Price: 106€
Duration: 2 hours Price: 184€

Fig. 3: Terms marked with hyperlinks

[5] As previously mentioned, the application is developed in the framework of an R&D project concerning health and beauty terminology and, thus, the terminological database that is provided by default in the system contains terms from this domain.

Along with this first step of uploading and processing a working text, it is important to customise users' preferences. In Inteliterm this option is provided in a specific tag of the program named Preferences (see Figure 4), which is accesible by following the path Edit > Preferences.



Information sources

Specialised resources: ☑ ES ☑ EN □ DE □ IT

☑ [EN] Geospa
☑ [EN] Solimar
☑ [EN] Health and fitness travel
☑ [EN] Spa break direct
☑ [EN] Body and soul
☑ [EN] Andalucia te quiere
☑ [EN] Aguas Salinas
☑ [EN] Spa finder
☑ [EN] One Travel
☑ [EN] European Spas
☑ [EN] Travel to wellness
□ [IT] Benessere florido
□ [IT] Hotel benessere
□ [IT] Glosari.it
□ [IT] Metafora Viaggi
□ [IT] Veneto Terme
□ [IT] Welcome to Italy
□ [IT] Whotels.it
□ [IT] Ayurveda italia
□ [IT] Federterme
□ [IT] Benessere viaggi
□ [IT] Glosario della bellezza
□ [IT] Terme.info
□ [IT] Angelini
□ [IT] Glossario dei termini
☑ [ES] Dic. salud, belleza y calidad de vida
☑ [ES] Dic. cosmética, bienestar y belleza
☑ [ES] Diccionario termal
☑ [ES] Termalismo
☑ [ES] Glosario Cosmetologia Actual
☑ [ES] poraqui.net
☑ [ES] Saludterapia
☑ [ES] beautymarket.es
☑ [ES] Nieves de Antonio
☑ [ES] elcuerpo.es
☑ [ES] AECC
☑ [ES] todopiscinas.es
☑ [ES] Eroski-consumer
☑ [ES] Termasworld
☑ [ES] Hotel-balneario La Hermida
□ [DE] Die Kur
□ [DE] mobile massage München
□ [DE] Amashi massage service
□ [DE] Wellnessguide-München
□ [DE] Glossar für Reise und Wellness
□ [DE] Reise im web
□ [DE] Wellness-ABC
□ [DE] Makeupbeauty.de
□ [DE] Urlaub in Bad Tölz
□ [DE] reisen.de

Entry fields

☑ Term (ES)
☑ Term (EN)
□ Term (DE)
□ Term (IT)
☑ Grammatical category
☑ Gender
☑ Grammatical number
☑ Domain
☑ Context
☑ Definition
☑ Collocations
☑ Remarks

*Extra

☑ Concordances
☑ Related terms
☑ Cognates
☑ ES ☑ EN □ DE □ IT
☑ N-grams N= 2
☑ Parallel texts

Fig. 4: Customising users' preferences

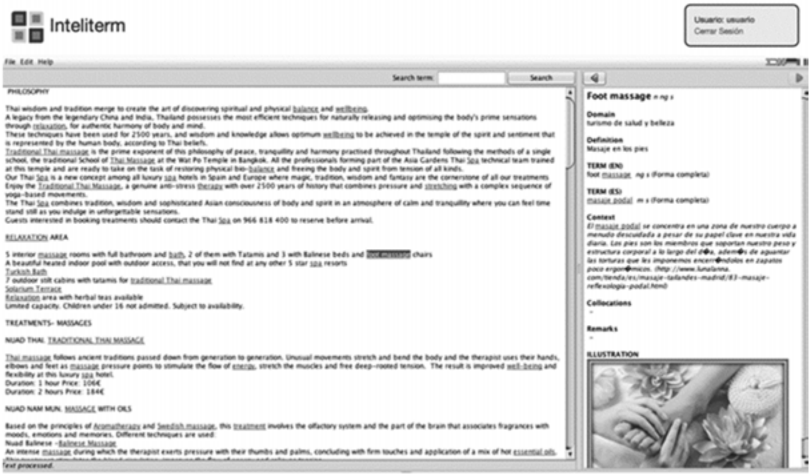
With this option, the user can select the information that will be displayed in the terminological entry, together with all the different functionalities that will be available in the tool (concordances, related terms, cognates, n-grams, and parallel texts) and the external resources and the languages that will be considered in carrying out external searches (see below).

In the following sections, we describe each of the three modules and their main features in more detail.

3.1 Term search

In order to aid access and efficient lookup for terminological information, this application offers four different types of terminology lookup to meet users' needs by minimising their effort when searching for terms.

Search in Context is one of the novel types of searches offered in the tool. It is performed by double-clicking on any of the hyperlinks marked in the working text (Figure 3 above). Once clicked, the information contained in the terminological database regarding the terms will be automatically displayed on the right-hand side of the window in the form of a terminological entry, including the information that has been selected from the *Preference* option in the previous step. Consequently, users can quickly and easily access the information contained in the terminological database without leaving the application or losing sight of the working text (Figure 5).



PHILOSOPHY

Thai wisdom and tradition merge to create the art of discovering spiritual and physical balance and wellbeing.
A legacy from the legendary China and India, Thailand possesses the most efficient techniques for naturally releasing and optimising the body's prime sensations through relaxation, for authentic harmony of body and mind.
These techniques have been used for 2500 years, and wisdom and knowledge allows optimum wellbeing to be achieved in the temple of the spirit and sentiment that is represented by the human body, according to Thai beliefs.
Traditional Thai massage is the prime exponent of this philosophy of peace, tranquillity and harmony practised throughout Thailand following the methods of a single school, the traditional School of Thai Massage at the Wat Po Temple in Bangkok. All the professionals forming part of the Asia Gardens Thai Spa technical team trained at this temple and are ready to take on the task of restoring physical bio-balance and freeing the body and spirit from tension of all kinds.
Our Thai Spa is a new concept among all luxury spa hotels in Spain and Europe where magic, tradition, wisdom and fantasy are the cornerstone of all our treatments
Enjoy the Traditional Thai Massage, a genuine anti-stress therapy with over 2500 years of history that combines pressure and stretching with a complex sequence of yoga-based movements.
The Thai Spa combines tradition, wisdom and sophisticated Asian consciousness of body and spirit in an atmosphere of calm and tranquility where you can feel time stand still as you indulge in unforgettable sensations.
Guests interested in booking treatments should contact the Thai Spa on 966 818 400 to reserve before arrival.

RELAXATION AREA

5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath
7 outdoor stilt cabins with tatamis for traditional Thai massage
Solarium Terrace
Relaxation area with herbal teas available
Limited capacity. Children under 16 not admitted. Subject to availability.

TREATMENTS- MASSAGES

NUAD THAI. TRADITIONAL THAI MASSAGE

Thai massage follows ancient traditions passed down from generation to generation. Unusual movements stretch and bend the body and the therapist uses their hands, elbows and feet as massage pressure points to stimulate the flow of energy, stretch the muscles and free deep-rooted tension. The result is improved well-being and flexibility at this luxury spa hotel.
Duration: 1 hour Price: 106€
Duration: 2 hours Price: 184€

NUAD NAM MUN. MASSAGE WITH OILS

Based on the principles of Aromatherapy and Swedish massage, this treatment involves the olfactory system and the part of the brain that associates fragrances with moods, emotions and memories. Different techniques are used:
Nuad Balinese –Balinese Massage
An intense massage during which the therapist exerts pressure with their thumbs and palms, concluding with firm touches and application of a mix of hot essential oils.

Text processed.

Foot massage *n ng s*

Domain
turismo de salud y belleza

Definition
Masaje en los pies

TERM (EN)
foot massage *ng s* (Forma completa)

TERM (ES)
masaje podal *m s* (Forma completa)

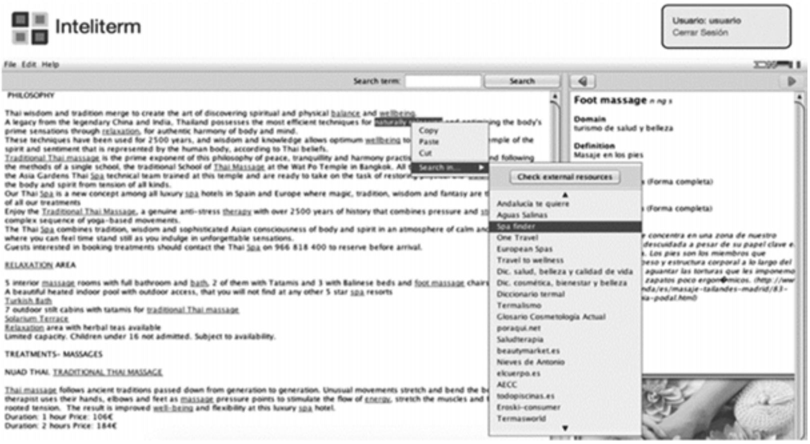
Context
El masaje podal se concentra en una zona de nuestro cuerpo a menudo descuidada a pesar de su papel clave en nuestra vida diaria. Los pies son los miembros que soportan nuestro peso y estructura corporal a lo largo del d�a, adem�s de aguantar las torturas que les imponemos encerr�ndolos en zapatos poco ergon�micos. (http://www.lunalanna.com/tienda/es/masaje-tailandes-madrid/83-masaje-reflexologia-podal.html)

Collocations
–

Remarks
–

ILLUSTRATION

Fig. 5: Terminological entry on the right side of the screen after clicking on a marked term



PHILOSOPHY

Thai wisdom and tradition merge to create the art of discovering spiritual and physical balance and wellbeing.
A legacy from the legendary China and India, Thailand possesses the most efficient techniques for naturally releasing and optimising the body's prime sensations through relaxation, for authentic harmony of body and mind.
These techniques have been used for 2500 years, and wisdom and knowledge allows optimum wellbeing to be achieved in the temple of the spirit and sentiment that is represented by the human body, according to Thai beliefs.
Traditional Thai massage is the prime exponent of this philosophy of peace, tranquillity and harmony practised throughout Thailand following the methods of a single school, the traditional School of Thai Massage at the Wat Po Temple in Bangkok. All the Asia Gardens Thai Spa technical team trained at this temple and are ready to take on the task of restoring the body and spirit from tension of all kinds.
Our Thai Spa is a new concept among all luxury spa hotels in Spain and Europe where magic, tradition, wisdom and fantasy are the of all our treatments
Enjoy the Traditional Thai Massage, a genuine anti-stress therapy with over 2500 years of history that combines pressure and complex sequence of yoga-based movements.
The Thai Spa combines tradition, wisdom and sophisticated Asian consciousness of body and spirit in an atmosphere of calm and where you can feel time stand still as you indulge in unforgettable sensations.
Guests interested in booking treatments should contact the Thai Spa on 966 818 400 to reserve before arrival.

RELAXATION AREA

5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath
7 outdoor stilt cabins with tatamis for traditional Thai massage
Solarium Terrace
Relaxation area with herbal teas available
Limited capacity. Children under 16 not admitted. Subject to availability.

TREATMENTS- MASSAGES

NUAD THAI. TRADITIONAL THAI MASSAGE

Thai massage follows ancient traditions passed down from generation to generation. Unusual movements stretch and bend the therapist uses their hands, elbows and feet as massage pressure points to stimulate the flow of energy, stretch the muscles and rooted tension. The result is improved well-being and flexibility at this luxury spa hotel.
Duration: 1 hour Price: 106€
Duration: 2 hours Price: 184€

Foot massage *n ng s*

Domain
turismo de salud y belleza

Definition
Masaje en los pies

(Forma completa)

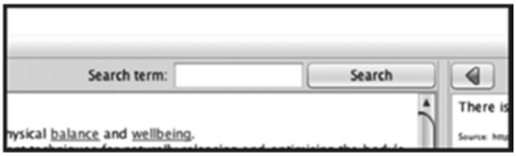
(Forma completa)

e concentra en una zona de nuestro descuidada a pesar de su papel clave e. Los pies son los miembros que peso y estructura corporal a lo largo del aguantar las torturas que les imponemo zapatos poco ergon�micos. (http://ww nda/es/masaje-tailandes-madrid/83-ia-podal.html)

Fig. 6: Steps to perform external search

A second type of search that is unique to this tool is the External Search, which consists of the possibility to launch queries to a wide range of external resources on the domain of health and beauty tourism (which are selected by the users in the *Preference* menu) that are linked to the application but that are external to it. To perform such searches just three simple steps are required: 1. Select the desired word or phrase with the cursor, 2. Click on the right mouse button, and 3. Select the desired external resource(s) from the provided list (an option to select all the available resources is also included).

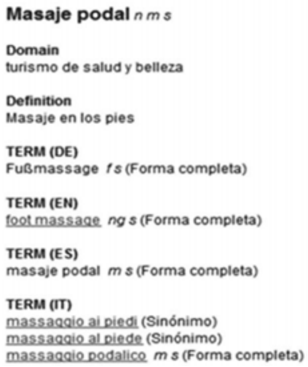
Apart from these two, the application also offers other two more conventional searches: traditional searches based on terms, which consist of inserting the desired term in the search field (Figure 7); and cross-reference searches, which are based on internal hyperlinks (Figure 8).



There is

ysical balance and wellbeing.

Fig. 7: Search by terms


Masaje podal *n m s*

Domain
turismo de salud y belleza

Definition
Masaje en los pies

TERM (DE)
Fußmassage *f s* (Forma completa)

TERM (EN)
foot massage *ng s* (Forma completa)

TERM (ES)
masaje podal *m s* (Forma completa)

TERM (IT)
massaggio ai piedi (Sinónimo)
massaggio al piede (Sinónimo)
massaggio podalico *m s* (Forma completa)

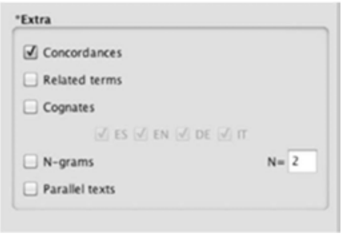
Fig. 8: Cross-reference searches

In short, the search options included in *Inteliterm* aim at speeding up translators' research tasks by saving time and effort and by making these searches more productive and efficient.

3.2 Corpus management tool

The benefits of corpus use in translation are manifold (cf. Corpas Pastor 2001, 2004, 2008, Beeby et al. 2009, Zanettin 2012, Corpas Pastor/Seghiri 2016). Both comparable and parallel corpora offer a range of possibilities to translators that are beyond other tools and resources. Corpora enable users to consult terms in real contexts and their right and left co-texts, to check their function in a sentence and the grammatical use, to confirm abbreviations and developed forms, and to study register and style, among other uses. In its aim of filling the existing gap and providing efficient and high quality results, *Inteliterm* provides tools to manage two diverse but complementary corpora dealing with health and beauty tourism: one is a comparable corpus composed of original texts in the working languages (English, Spanish, German and Italian), and the other one is a parallel corpus comprising original texts and their translations.

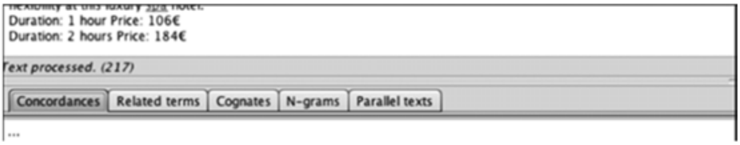
In line with our main purpose of customisation and user-friendliness, users can select or deselect the different options offered by the application concerning corpus management in the *Preference screen* (Figure 9) according to their needs.



*Extra

☑ Concordances
☐ Related terms
☐ Cognates
☑ ES ☑ EN ☑ DE ☑ IT
☐ N-grams N= 2
☐ Parallel texts

Fig. 9: Options for corpus management on the bottom right corner of the Preference screen



Duration: 1 hour Price: 106€
Duration: 2 hours Price: 184€

Text processed. (217)

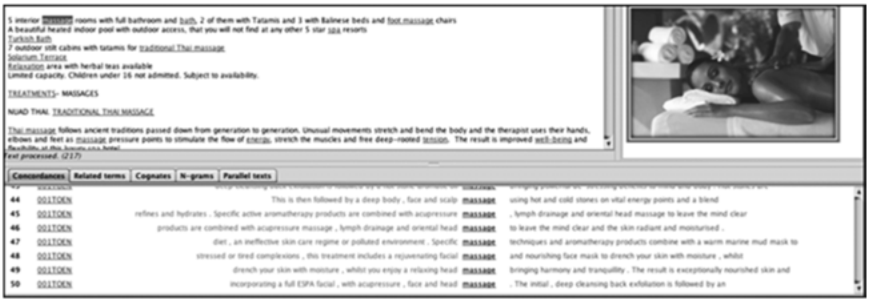
Fig. 10: Options for corpus management on the bottom of the main screen

The main uses of corpora for translation are included in the application:

- *Concordance searches*: to search for real uses of terms and co-texts (see Figure 11).
- *Related terms*: to find derivatives and other formal variations (see Figure 12).
- *Cognates*: to detect equivalents that share a similar form and a common root. Users can also select their working languages.

- *N-grams*: to extract terminological units formed by the selected term and *n* units[6] (see Figure 13).
- *Parallel texts*: to consult the parallel texts where the term in question is found (see Figure 14).

To launch these searches, users only need to double-click on the terms they desire to consult in the working text and select the corresponding tabs in the main screen. The results are automatically displayed, together with some data regarding frequency of occurrences in the corpora.



5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath
7 outdoor stilt cabins with tatamis for traditional Thai massage
Solarium Terrace
Relaxation area with herbal teas available
Limited capacity. Children under 16 not admitted. Subject to availability.

TREATMENTS- MASSAGES

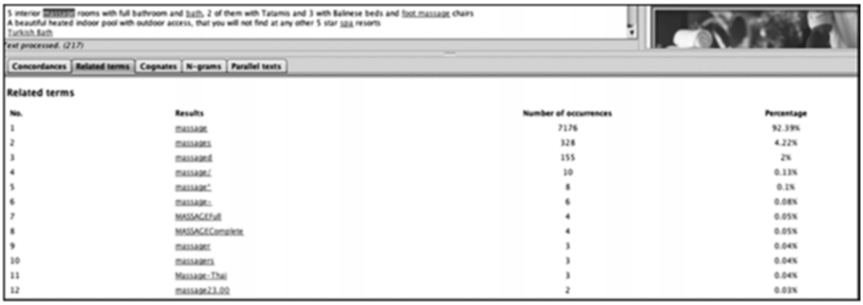
NUAD THAI. TRADITIONAL THAI MASSAGE

Thai massage follows ancient traditions passed down from generation to generation. Unusual movements stretch and bend the body and the therapist uses their hands, elbows and feet as massage pressure points to stimulate the flow of energy, stretch the muscles and free deep-rooted tension. The result is improved well-being and

Text processed. (217)

44	001TOEN	This is then followed by a deep body , face and scalp	massage	using hot and cold stones on vital energy points and a blend
45	001TOEN	refines and hydrates . Specific active aromatherapy products are combined with acupressure	massage	, lymph drainage and oriental head massage to leave the mind clear
46	001TOEN	products are combined with acupressure massage , lymph drainage and oriental head	massage	to leave the mind clear and the skin radiant and moisturised .
47	001TOEN	diet , an ineffective skin care regime or polluted environment . Specific	massage	techniques and aromatherapy products combine with a warm marine mud mask to
48	001TOEN	stressed or tired complexions , this treatment includes a rejuvenating facial	massage	and nourishing face mask to drench your skin with moisture , whilst
49	001TOEN	drench your skin with moisture , whilst you enjoy a relaxing head	massage	bringing harmony and tranquillity . The result is exceptionally nourished skin and
50	001TOEN	incorporating a full ESPA facial , with acupressure , face and head	massage	. The initial , deep cleansing back exfoliation is followed by an

Fig. 11: Concordances for the search “massage”



5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath

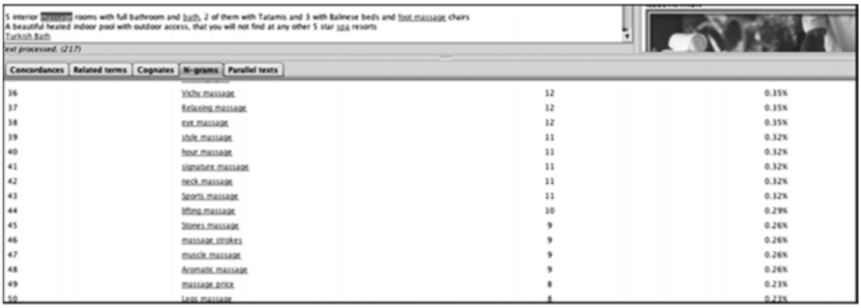
Text processed. (217)

Related terms

No.	Results	Number of occurrences	Percentage
1	massage	7176	92.39%
2	massages	328	4.22%
3	massaged	155	2%
4	massage/	10	0.13%
5	massage"	8	0.1%
6	massage-	6	0.08%
7	MASSAGEFull	4	0.05%
8	MASSAGEComplete	4	0.05%
9	massager	3	0.04%
10	massagers	3	0.04%
11	Massage-Thai	3	0.04%
12	massage23.00	2	0.03%

Fig. 12: Related terms for the search “massage”

[6] The *n* value can be also selected in the *Preference* screen.

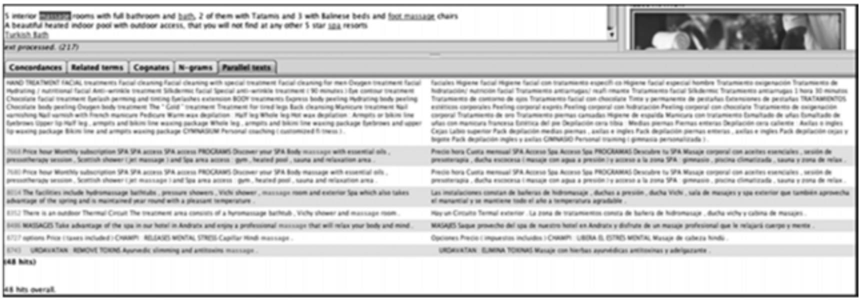


5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath

ext processed. (217)

36	Vichy massage	12	0.35%
37	Relaxing massage	12	0.35%
38	eye massage	12	0.35%
39	style massage	11	0.32%
40	hour massage	11	0.32%
41	signature massage	11	0.32%
42	neck massage	11	0.32%
43	Sports massage	11	0.32%
44	lifting massage	10	0.29%
45	Stones massage	9	0.26%
46	massage strokes	9	0.26%
47	muscle massage	9	0.26%
48	Aromatic massage	9	0.26%
49	massage price	8	0.23%
50	Legs massage	8	0.23%

Fig. 13: N-grams for the search "massage"



5 interior massage rooms with full bathroom and bath, 2 of them with Tatamis and 3 with Balinese beds and foot massage chairs
A beautiful heated indoor pool with outdoor access, that you will not find at any other 5 star spa resorts
Turkish Bath

ext processed. (217)

HAND TREATMENT FACIAL treatments Facial cleaning Facial cleaning with special treatment Facial cleaning for men Oxygen treatment facial Hydrating / nutritional facial Anti-wrinkle treatment Silkdermic facial Special anti-wrinkle treatment (90 minutes) Eye contour treatment Chocolate facial treatment Eyelash perming and tinting Eyelashes extension BODY treatments Express body peeling Hydrating body peeling Chocolate body peeling Oxygen body treatment The " Gold " treatment Treatment for tired legs Back cleansing Manicure treatment Nail varnishing Nail varnish with French manicure Pedicure Warm wax depilation : Half leg Whole leg Hot wax depilation : Armpits or bikini line Eyebrows Upper lip Half leg , armpits and bikini line waxing package Whole leg , armpits and bikini line waxing package Eyebrows and upper lip waxing package Bikini line and armpits waxing package GYMNASIUM Personal coaching (customized fi tness) .	faciales Higiene facial Higiene facial con tratamiento específi co Higiene facial especial hombre Tratamiento oxigenación Tratamiento de hidratación/ nutrición facial Tratamiento antiarrugas/ reafi rmante Tratamiento facial Silkdermic Tratamiento antiarrugas 1 hora 30 minutos Tratamiento de contorno de ojos Tratamiento facial con chocolate Tinte y permanente de pestañas Extensiones de pestañas TRATAMIENTOS estéticos corporales Peeling corporal exprés Peeling corporal con hidratación Peeling corporal con chocolate Tratamiento de oxigenación corporal Tratamiento de oro Tratamiento piernas cansadas Higiene de espalda Manicura con tratamiento Esmaltado de uñas Esmaltado de uñas con manicura francesa Estética del pie Depilación cera tibia : Medias piernas Piernas enteras Depilación cera caliente : Axilas o ingles Cejas Labio superior Pack depilación medias piernas , axilas e ingles Pack depilación piernas enteras , axilas e ingles Pack depilación cejas y bigote Pack depilación ingles y axilas GIMNASIO Personal training (gimnasia personalizada) .
7668 Price hour Monthly subscription SPA SPA access SPA access PROGRAMS Discover your SPA Body massage with essential oils , pressotherapy session , Scottish shower (jet massage) and Spa area access : gym , heated pool , sauna and relaxation area .	Precio hora Cuota mensual SPA Acceso Spa Acceso Spa PROGRAMAS Descubre tu SPA Masaje corporal con aceites esenciales , sesión de presoterapia , ducha escocesa (masaje con agua a presión) y acceso a la zona SPA : gimnasio , piscina climatizada , sauna y zona de relax .
7680 Price hour Monthly subscription SPA SPA access SPA access PROGRAMS Discover your SPA Body massage with essential oils , pressotherapy session , Scottish shower (jet massage) and Spa area access : gym , heated pool , sauna and relaxation area .	Precio hora Cuota mensual SPA Acceso Spa Acceso Spa PROGRAMAS Descubre tu SPA Masaje corporal con aceites esenciales , sesión de presoterapia , ducha escocesa (masaje con agua a presión) y acceso a la zona SPA : gimnasio , piscina climatizada , sauna y zona de relax .
8014 The facilities include hydromassage bathtubs , pressure showers , Vichi shower , massage room and exterior Spa which also takes advantage of the spring and is maintained year round with a pleasant temperature .	Las instalaciones constan de bañeras de hidromasaje , duchas a presión , ducha Vichi , sala de masajes y spa exterior que también aprovecha el manantial y se mantiene todo el año a temperatura agradable .
8352 There is an outdoor Thermal Circuit The treatment area consists of a hyromassage bathtub , Vichy shower and massage room .	Hay un Circuito Termal exterior . La zona de tratamientos consta de bañera de hidromasaje , ducha vichy y cabina de masajes .
8486 MASSAGES Take advantage of the spa in our hotel in Andratx and enjoy a professional massage that will relax your body and mind .	MASAJES Saque provecho del spa de nuestro hotel en Andratx y disfrute de un masaje profesional que le relajará cuerpo y mente .
8727 options Price (taxes included) CHAMPI : RELEASES MENTAL STRESS Capillar Hindi massage .	Opciones Precio (impuestos incluidos) CHAMPI : LIBERA EL ESTRÉS MENTAL Masaje de cabeza hindú .
8743 URDAVATAN : REMOVE TOXINS Ayurvedic slimming and antitoxins massage .	URDAVATAN : ELIMINA TOXINAS Masaje con hierbas ayurvédicas antitoxinas y adelgazante .

(48 hits)

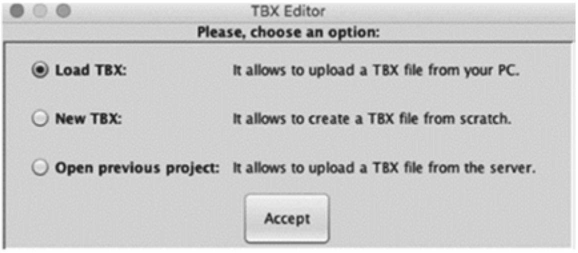
48 hits overall.

Fig. 14: Parallel texts for the search "massage"

3.3 Terminology Management Tool (TMT)

Terminological resources, either of high or low quality, electronic or paper-based, are mostly static and unidirectional. Users cannot interact or modify the searches offered, nor the content, and they are obliged to use different applications to create their own resources that are sometimes incompatible with one another. By contrast, *Inteliterm* has been designed for efficient and user-friendly use. It helps reduce unproductive searches and efforts when learning how to use applications and/or employing several applications simultaneously. Thus, this application includes by default a database about health and beauty tourism in the working languages (English, Spanish, German and Italian), but it also offers the possibility to both enlarge this database by adding more terms and create or edit users' own databases according to their needs.

The databases can be imported or exported in the standard format TBX (TermBase eXchange) (ISO 30042: 2008), which greatly enhances the compatibility and exchange of databases among users as well as the interaction with other TBX-based applications.



Please, choose an option:

● Load TBX: It allows to upload a TBX file from your PC.

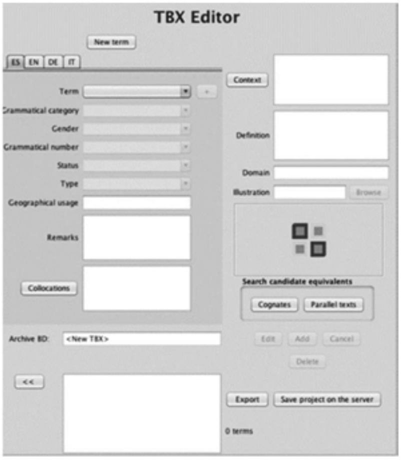
○ New TBX: It allows to create a TBX file from scratch.

○ Open previous project: It allows to upload a TBX file from the server.

Fig. 15: Working with the TBX Editor

Once this TMT is selected from the main page of *Inteliterm* (Edit > Editor), the first decision to make is the preferred option to start: 1. *Load TBX*, if there exists an external TBX database that can be imported in the tool; 2. *New TBX*, if a new TBX database is to be created; and 3. *Open previous project*, if the user is looking to continue working on a previous project (Figure 15). The next step is either to upload an external database, to select the previous project or to create and name a new one.
In Figure 15, the term entry for each term is displayed. As observed, the editor interface is very intuitive and easy to use. By clicking on the *New term* button at the top, all the fields are unblocked and users can start adding the relevant information in the different working languages of their projects (selecting them from the given languages: English, Spanish, German and Italian). The fields that are given are those considered essential for a terminological resource aimed at translators (Durán-Muñoz 2010, 2012)[7], though none of them are compulsory, apart from the *Term* field. Therefore, here users can also choose the information they enter in the database.

[7] At this point, it is worth recalling the possibility given to users on the Preference screen to select/deselect the fields they want to have displayed in the terminological entries when using Inteliterm during their translation task.



Archive BD: <New TBX>

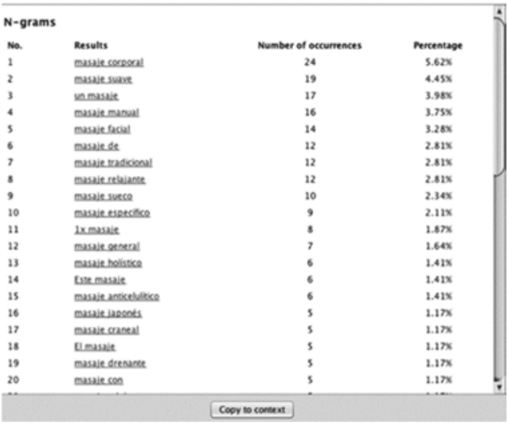
0 terms

Fig. 16: Term entry in the *TBX Editor*

When editing a database, the number of terms comprising the database is shown at the bottom. In Figure 16, the number is zero, since the database is a new creation.

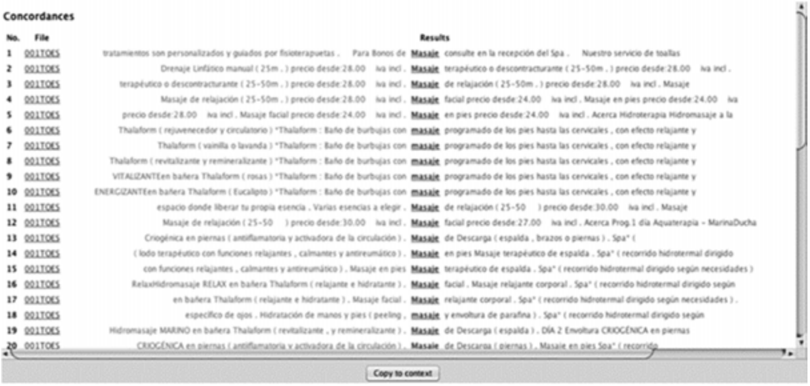
Among these fields, there are some that merit further description: *Collocations*, *Context*, *Illustration*, and the options to *Search candidate equivalents*.

Collocations and *Context* give the opportunity to query the comparable corpus linked to the editor. By clicking on their respective buttons, a new screen displaying the information related to the term in question is provided, either in the form of a list of n-grams (Figure 17) or concordance lines (Figure 18). Users only need to select the correct results and the information is automatically added to the corresponding field.


N-grams

No.	Results	Number of occurrences	Percentage
1	masaje corporal	24	5.62%
2	masaje suave	19	4.45%
3	un masaje	17	3.98%
4	masaje manual	16	3.75%
5	masaje facial	14	3.28%
6	masaje de	12	2.81%
7	masaje tradicional	12	2.81%
8	masaje relajante	12	2.81%
9	masaje sueco	10	2.34%
10	masaje especifico	9	2.11%
11	1x masaje	8	1.87%
12	masaje general	7	1.64%
13	masaje holistico	6	1.41%
14	Este masaje	6	1.41%
15	masaje anticelulitico	6	1.41%
16	masaje japonés	5	1.17%
17	masaje craneal	5	1.17%
18	El masaje	5	1.17%
19	masaje drenante	5	1.17%
20	masaje con	5	1.17%

Fig. 17: N-grams for the term “masaje”



Concordances

No.	File		Results	
1	001TOES	tratamientos son personalizados y guiados por fisioterapuetas . Para Bonos de	Masaje	consulte en la recepción del Spa . Nuestro servicio de toallas
2	001TOES	Drenaje Linfático manual (25m .) precio desde:28.00 iva incl .	Masaje	terapéutico o descontracturante (25-50m .) precio desde:28.00 iva incl .
3	001TOES	terapéutico o descontracturante (25-50m .) precio desde:28.00 iva incl .	Masaje	de relajación (25-50m .) precio desde:28.00 iva incl . Masaje
4	001TOES	Masaje de relajación (25-50m .) precio desde:28.00 iva incl .	Masaje	facial precio desde:24.00 iva incl . Masaje en pies precio desde:24.00 iva
5	001TOES	precio desde:28.00 iva incl . Masaje facial precio desde:24.00 iva incl .	Masaje	en pies precio desde:24.00 iva incl . Acerca Hidroterapia Hidromasaje a la
6	001TOES	Thalaform (rejuvenecedor y circulatorio) *Thalaform : Baño de burbujas con	masaje	programado de los pies hasta las cervicales , con efecto relajante y
7	001TOES	Thalaform (vainilla o lavanda) *Thalaform : Baño de burbujas con	masaje	programado de los pies hasta las cervicales , con efecto relajante y
8	001TOES	Thalaform (revitalizante y remineralizante) *Thalaform : Baño de burbujas con	masaje	programado de los pies hasta las cervicales , con efecto relajante y
9	001TOES	VITALIZANTEen bañera Thalaform (rosas) *Thalaform : Baño de burbujas con	masaje	programado de los pies hasta las cervicales , con efecto relajante y
10	001TOES	ENERGIZANTEen bañera Thalaform (Eucalipto) *Thalaform : Baño de burbujas con	masaje	programado de los pies hasta las cervicales , con efecto relajante y
11	001TOES	espacio donde liberar tu propia esencia . Varias esencias a elegir .	Masaje	de relajación (25-50) precio desde:30.00 iva incl . Masaje
12	001TOES	Masaje de relajación (25-50) precio desde:30.00 iva incl .	Masaje	facial precio desde:27.00 iva incl . Acerca Prog.1 día Aquaterapia – MarinaDucha
13	001TOES	Criogénica en piernas (antiflamatoria y activadora de la circulación) .	Masaje	de Descarga (espalda , brazos o piernas) . Spa* (
14	001TOES	(lodo terapéutico con funciones relajantes , calmantes y antireumático) .	Masaje	en pies Masaje terapéutico de espalda . Spa* (recorrido hidrotermal dirigido
15	001TOES	con funciones relajantes , calmantes y antireumático) . Masaje en pies	Masaje	terapéutico de espalda . Spa* (recorrido hidrotermal dirigido según necesidades)
16	001TOES	RelaxHidromasaje RELAX en bañera Thalaform (relajante e hidratante) .	Masaje	facial . Masaje relajante corporal . Spa* (recorrido hidrotermal dirigido según
17	001TOES	en bañera Thalaform (relajante e hidratante) . Masaje facial .	Masaje	relajante corporal . Spa* (recorrido hidrotermal dirigido según necesidades) .
18	001TOES	específico de ojos . Hidratación de manos y pies (peeling ,	masaje	y envoltura de parafina) . Spa* (recorrido hidrotermal dirigido según
19	001TOES	Hidromasaje MARINO en bañera Thalaform (revitalizante , y remineralizante) .	Masaje	de Descarga (espalda) . DÍA 2 Envoltura CRIOGÉNICA en piernas
20	001TOES	CRIOGÉNICA en piernas (antiflamatoria y activadora de la circulación) .	Masaje	de Descarga (piernas) . Masaje en pies Spa* (recorrido

Fig. 18: Concordance lines for the term "masaje"

These options enable the combination of corpus and terminological management tools and bring together the advantages of both, increasing the quality of the final results by employing real texts and reducing users' efforts by implementing Natural Language Processing tools.

The *Search candidate* equivalents options, instead of querying the comparable corpus, offer the possibility to consult parallel corpora, i.e. original texts and their translations. *Cognates* and *Parallel texts* provide users with access to already translated material and help them find equivalents and, if needed, pragmatic information for the *Remarks* field. Users only need to click on the respective buttons to obtain information regarding these options either in the form of a list or bilingual concordance lines.

Finally, the *Illustration* field allows users to upload an image to complete the description of the term being investigated. Many scholars highlight the advantages of adding descriptive images to term entries (Prieto Velasco/Tercedor Sánchez 2014, Tercedor Sánchez/López Rodríguez 2012, among others), since they clarify concepts and help to better grasp definitions. The Inteliterm TBX Editor, bearing this in mind, provides this possibility by just clicking on the Browse option and selecting the desired image.

When the term entry is completed, users will add it to their database and it will be included in the blank field at the bottom of the screen, where it will be always available to be edited or deleted in the future. As a final step, users are also able to export the terminological database in .tbx format so that it can be used in other applications, such as translation memories or other terminology management tools. In a similar way, users can also select their databases to be queried in the *Inteliterm* searches and, thus, substitute the default database included in the system for a new one dealing with the same or other specialised domain.

4 User evaluation and feedback

To prove the functionalities of *Inteliterm* and learn about users' satisfaction regarding the tool, we conducted an empirical experiment with students of Translation and Interpreting at several Spanish universities (University of Málaga, University of Alcalá and University of Castilla-La Mancha) during the second semester of the academic year 2016-17. The total number of participants was 167.

Sessions were organised as showcases in order to explain the functionalities of the tool, and time was allotted for the participants to explore the tool by themselves and to give their opinions by means of a questionnaire (Annex 1).

In general, it can be stated that the results regarding users' satisfaction were very positive, since participants conveyed their satisfaction and willingness to use *Inteliterm* in all sessions. The idea of developing *Inteliterm* as a web-based tool was also welcomed by the participants, since the response "online dictionaries" were around 80% in all the sessions. This proves that students are more used to employing web resources than stand-alone applications, paper-based dictionaries or other resources.

Concerning the functionality of the tool, most participants considered that access to the tool was easy (53.85%) or very easy (3.85%) and only 7.69% considered it very difficult, because of the adjustments that are required by Java when using the tool for the first time. As for ease of use, most of the participants indicated that it is easy and intuitive to use (65%).

Question 7 referred to the usefulness of *Inteliterm* in their daily life as a translator, and more than 70% of participants considered it useful (58%) or very useful (15%), indicating they would use it for their translation tasks if it were available.

Participants were also asked to answer an open question about what they would add to the tool and what they would change. Their comments are essential for deciding on future improvements. Some of the most frequent comments and suggestions are provided below:

- The need to expand the number of specialised domains included in the application and, thus, the terminological databases related to these new domains.
- The possibility for users to upload their own comparable or parallel corpus from other domains to manage them with the tool.
- Easier access through other browsers and avoiding requiring adjustments by Java.
- Making the User Help option more visible to users as well as the different steps required to customise the tool.
- Keeping external resources updated.

Finally, participants were also asked to rate their satisfaction with the tool on a scale of 1 to 10, 10 being the highest score. Judging from their answers, there is a high level of satisfaction with the tool, since more than 75% of participants graded *Inteliterm* as 7 or 8. The average mark was 7.5.

5 Conclusion

Inteliterm fills an existing gap in the array of resources aimed at translators. This next-generation tool helps reduce the number of unproductive terminological searches by means of contextual searches, cross searches and external searches provided. This novel LSP lexicographic tool takes into account translators' needs and provides several types of consultations, easy access to terminological information and customisation options. Additionally, it offers the possibility to create, edit and exchange users' own terminological databases in a very easy way that can likewise be used for consultation. All these features, combined with a corpus management tool, bring many advantages to translation and terminological work, both with comparable and parallel corpora, and NLP applications. *Inteliterm* proves to be a dynamic, flexible and proactive resource that facilitates the documentary work of translators and, as a result, improves efficiency and results of the final product.

Further improvements will take into account the feedback given by the participants in the experiments a well as the necessary requirement to update external resources related to health and beauty tourism and other tourism segments. Finally, we will explore the possibility of adding a terminology extractor to the TBX Editor.

6 Bibliography

Agirre, Eneko/Arregi, Xabier/Artola, Xabier/Díaz De Ilarraza, Arantza/Sarasola, Kepa/Soroa, Aitor (2003): An Intelligent Dictionary Help System. Encyclopedia of Library and Information Sciences. New York: Taylor & Francis.

Arnaudov, Todor/Mitkov, Ruslan (2008): Smarty - Extendable Framework for Bilingual and Multilingual Comprehension Assistants. Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC'08. Marrakech: European Language Resources Association, 3287–3292.

Beeby, Allison/Rodríguez Inés, Patricia/Sánchez-Gijón, Pilar (2009): Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate. Amsterdam: John Benjamins.

Bergenholtz, Henning (2011): Access to and Presentation of Needs-adapted Data in Monofunctional Internet Dictionaries. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography. The Internet, Digital Initiatives and Lexicography. London, New York: Continuum, 30–53.

Bowker, Lynne/Corpas Pastor, Gloria (2015): Translation Technology. In: Mitkov, Ruslan (ed.): The Oxford Handbook of Computational Linguistics. 2nd edition. Oxford: Oxford University Press. [Online first publication].

Corpas Pastor, Gloria (2001): Compilación de un corpus ad hoc para la enseñanza de la traducción inversa. In: TRANS 5. 155–179.

Corpas Pastor, Gloria (2008): Investigar con corpus en traducción: los retos de un nuevo paradigma. Frankfurt am Main: Peter Lang.

Corpas Pastor, Gloria/Seghiri, Míriam (eds.) (2016): Corpus-based Approaches to Translation and Interpreting. From Theory to Applications. Frankfurt am Main: Peter Lang.

De Schryver, Gilles-Maurice (2009): State-of-the-Art Software to Support Intelligent Lexicography. In: Zhu, R. (ed.): Proceedings of the International Seminar on Kangxi Dictionary & Lexicology. Beijing: Beijing Normal University, 565–580.

Durán-Muñoz, Isabel (2010): Specialized lexicographical resources: a survey of translators' needs. In: Granger, Sylviane/Paquot, Magali (eds.): eLexicography in the 21st century: New Challenges, new applications. Proceedings of ELEX2009. Cahiers du Cental, vol. 7. Lovaine-La-Neuve: Presses Universitaires de Louvain, 55–66.

Durán-Muñoz, Isabel (2012): La ontoterminografía aplicada a la traducción: Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores. Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation 80. Frankfurt, New York: Peter Lang.

Durán-Muñoz, Isabel (2014): Nuevas posibilidades de búsqueda terminológica eficiente para los traductores: la herramienta Trandix. In: Vargas Sierra, C. (ed.): TIC, trabajo colaborativo e interacción en Terminología y Traducción. Granada: Editorial Comares, 201–212.

Durán-Muñoz, Isabel/Fernández Sola, Alejandro (2014): Trandix: a proactive tool to terminological searches by translators and its assessment. In: Calvo Rigual, Cesáreo/Calvi, Maria Vittoria (eds.): MONTI. Monografías de Traducción e Interpretación 6. Valencia: Universitat de Valéncia, 115–139.

Feldweg, Helmut/Breidt, Elisabeth (1996): COMPASS. An Intelligent Dictionary System for Reading Text in a Foreign Language. In: Kiefer, Ferenc/Kiss, Gábor/Pajzs, Júlia (eds.): Papers in Computational Lexicography. COMPLEX '96. Budapest: editorial, 53–62.

Nielsen, Sandro (2018): LSP Lexicography and Typology of Specialized Dictionaries. In Humbley, John/Budin, Gerhard/Laurén, Christer (eds.): Languages for Special Purposes: An International Handbook. Berlin, Boston: De Gruyter Mouton, 71–95.

Prieto Velasco, Juan Antonio/Tercedor Sánchez, Maribel (2014): The embodied nature of medical concepts: image schemas and language for pain. In: Cognitive processing 15(3), 283–296.

Prószéky, Gábor/Földes, András (2005): Between Understanding and Translating: A Context-Sensitive Comprehension Tool. In: Archive of Control Sciences 15 (4), 637–644.

Spohr, Dennis (2011): A Multi-layer Architecture for "Pluri-monofunctional" Dictionaries. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography. The Internet, Digital Initiatives and Lexicography. London, New York: Continuum, 103–120.

Tarp, Sven (2011): Lexicographical and Other e-Tools for Consultation Purposes: Towards the Individualization of Needs satisfaction. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography. The Internet, Digital Initiatives and Lexicography. London, New York: Continuum, 54–70.

Tercedor Sánchez, Maribel/López Rodríguez, Clara Inés (2012): Access to health in an intercultural setting: the role of corpora and images in grasping term variation. In: Linguistica Antverpiensia, New Series-Themes in Translation Studies (11), 247–268.

Whitelock, Pete/Edmonds, Philip (2000): The Sharp Intelligent Dictionary. In: Heid, Ulrich/Evert, Stefan/Lehmann, Egbert/Rohrer, Christian (eds.): Proceedings of the Ninth EURALEX International Congress (EURALEX 2000. Stuttgart: Universität Stuttgart, 871–876.

Zanettin, Federico (2012): Translation-driven corpora: Corpus resources for descriptive and applied translation studies. Manchester: St. Jerome Publishing.

Željko, Miran (2009): Improvements of Dictionaries – Suggestions by Evroterm. Proceedings of the INFuture2009: Digital Resources and Knowledge Sharing, Zagreb, 269–278.

ANNEX: Questionnaire on *Inteliterm*

1 Al traducir un texto de tema general, ¿qué tipo de herramientas sueles utilizar?

[What types of tools do you normally use when translating a general text?]

□ Diccionario en papel □ Diccionario electrónico □ Diccionario on-line
[paper dictionary] [e-dictionary] [on line dictionary]
□ Herramienta especializada (SDL MultiTerm, etc.) □ Otro: __________
[term tool (SDL MultiTerm, etc.] [other]

2 ¿Y de tema especializado?

[What types of tools do you normally use when translating a specialised text?]

□ Diccionario en papel □ Diccionario electrónico □ Diccionario on-line
[paper dictionary] [e-dictionary] [on line dictionary]
□ Herramienta especializada (SDL MultiTerm, etc.) □ Otro: __________
[term tool (SDL MultiTerm, etc.] [other]

3 ¿Con qué frecuencia usas herramientas de traducción asistida?

[How frequently do you use CAT tools?]

□ Siempre □ Muchas veces □ A veces □ Casi nunca □ Nunca
[Always] [Very frequently] [Sometimes] [Rarely] [Never]
Indica cuál: ____
[Which CAT tools do you use?]

4 Con respecto a lo que has visto de la herramienta *Inteliterm*, ¿crees que es fácil acceder a ella?

[After trying out *Inteliterm*, do you think it is easily accessible and user-friendly?]

□ Muy fácil □ Fácil □ Un poco complicado □ Muy difícil
[Very easy] [Easy] [A bit complicated, difficult] [Very difficult]
Comentario: ____
[Comments]

5 ¿Crees que es fácil de manejar?

[Do you find *Inteliterm* user-friendly?]

□ Muy fácil □ Fácil □ Un poco complicado □ Muy difícil
[Very easy] [Easy] [A bit complicated, difficult] [Very difficult]
Comentario: ____
[Comments]

6 ¿Crees que sería útil?

[Do you think *Inteliterm* could be helpful?]

□ Muy útil	□ Útil	□ Algo útil	□ Nada útil
[Very helpful]	[Hepful]	[Somewhat helpful]	[Not helpful at all]

Comentario: ____

[Comments]

7 ¿Qué añadirías a la herramienta?

[What else would you add to this tool?

8 ¿Qué cambiarías/quitarías de la herramienta?

[What would you change/remove from this tool?]

9 Valora del 1 al 10 la utilidad de la aplicación *Inteliterm*: ______

[Please rate the usefulness of *Inteliterm* on a scale from 1 to 10 (being 1 the lowest score)]

María José Domínguez Vázquez and Carlos Valcárcel Riveiro

PORTLEX as a multilingual and cross-lingual online dictionary

Abstract: This article focuses on the description of the fundamental aspects of PORTLEX, a valency oriented, multi/cross-lingual lexicographic resource focused on the noun phrase. Although special attention is paid to significant innovations related to its valency, cross-lingual and collaborative features, the different phases of the project that gave rise to PORTLEX, as well as the general outlines of its macro and microstructure, and database architecture are also explained. Beyond all technical issues, the often neglected matter of managing the work and responsibilities of the people involved first in the development and now in the regular updating of PORTLEX has been included as well. In this respect, particular emphasis is placed on the role of users and their channels of collaboration with the project within a community of practice also involving the editing team. The text concludes with some final reflections on the possible optimization of this tool by combining it with other existing lexicographic resources.

Keywords: multilingual online dictionary, cross-lingual dictionary, semi-collaborative dictionary, valency tool

1 Introduction

PORTLEX is mainly a lexicographical online tool that compiles multilingual data on the valency of the noun phrase in German, Galician, Spanish, Italian, and

María José Domínguez Vázquez: Universidade de Santiago de Compostela, Facultade de Filoloxía, Dpto. Filoloxía Inglesa e Alemá, Avda. de Castelao s/n, 15782 Santiago de Comostela (Spain), tel. +34 881811000 ext. 11761, majo.dominguez@usc.es
Carlos Valcárcel Riveiro: Universidade de Vigo, Facultade de Ciencias da Educación e do Deporte, Dpto. Filoloxía Inglesa, Francesa e Alemá, Campus da Xunqueira s/n, 36005 Pontevedra (Spain), tel. +34 986801732, carlos.valcarcel@uvigo.es

https://doi.org/10.1515/9783110607659-008

French[1]. This database has been developed since 2012 by an international and multidisciplinary team involving several institutions from Spain, Germany and Portugal: the University of Santiago de Compostela (Spain) as leading partner, the University of Vigo (Spain), the University of A Coruña (Spain), the Universität Nürnberg-Erlangen (Germany), Institut für Deutsche Sprache (Germany), the University of Seville (Spain), the University of Leipzig (Germany) and the Polytechnic Institute of Cávado and Ave (Portugal), who handled the programming work. The project was financed for three years (2012–2015) by the Spanish Ministry of Economy and Competitiveness, which allowed the development and hosting of the database.

PORTLEX has been developed as an accessible consultation tool for a wide range of users with different lexicographic needs. Its applications range from language learning and teaching to translation, grammar research, or natural language processing (NLP). Currently, PORTLEX is still being updated according to its initial goals:

- The development and maintaining of a multilingual annotated database based on valency theory and corpus linguistics. For the five languages mentioned above, PORTLEX stores and relates human-generated data on the valency of the noun phrase and its combinatory.
- The development and updating of a lexicographical user interface, that is to provide easy access to the database using a cross-lingual online dictionary format.
- The establishing of a virtual community of users and editors around PORTLEX. This is mostly achieved through an interface allowing the creation of different user profiles, from administrators and editors to learners and primary users.

This article will present the main features of PORTLEX, focusing mainly on its collaborative, cross-lingual and valency nature (section 2). Furthermore, its development stages (section 3) and its theoretical background (section 4) will be explained. Particular attention will be paid to the innovations implemented in PORTLEX (section 5) and to the structure of its database and its user interface, that is, the macro and microstructure of the online dictionary (section 6). Finally, the different types of

[1] The results of this research are related to the research projects “Portal Lexicográfico: Diccionario online modular multilingüe y corpus informatizado anotado de la frase nominal” (Ministry of Economy and Competitiveness, FFI2012-32456), “Generación multilingüe de estructuras argumentales del sustantivo y automatización de extracción de datos sintáctico-semánticos” (BBVA Foundation Grants for Scientific Research Teams 2017) and to the research project “Multilingual generator of noun argument structures with application in foreign language production” (Ministry of Economy, Industry and Competitiveness/Spanish State Research Agency, FFI2017-82454-P, and European Regional Development Fund). The present article has been undertaken within the framework of the RELEX network (ED431D R2016/046), financed by the Galician government.

searches will also be described (section 7) before some concluding reflections on the challenges ahead (section 8).

2 What is PORTLEX?

PORTLEX is not only an online reference dictionary, but also an annotated database and a community of users and editors. Thus, since its launch, this tool has gradually become a virtual meeting place for specialists, professionals, and students concerned with the noun valency from a multilingual point of view. The main characteristics of this tool are presented below, highlighting its electronic, collaborative, valential and multilingual features.

2.1 An electronic collaborative dictionary

According to the classification of Klosa (2013) and De Schryver (2003), PORTLEX is an online dictionary, since its users worldwide operate with different devices to access a database stored in an online server. Although users must request to open a profile through an online form, access to data is open and free. Furthermore, following the works of Abel and Meyer (2013) or Melchior (2014), PORTLEX could also be defined as a semi-collaborative dictionary, since users can add content to it, although under the supervision of the editing team. Users can collaborate directly, editing entries when they are given permission, and indirectly, sending information to the editor team by different means. A more detailed discussion of these issues will be presented in section 5.3.

Regarding its medial features, this dictionary was developed as an online and continuously updated resource based on hypertextualization, user interaction and combined access. For the time being, there are no plans to include multimedia content such as images, audio, or videos. Recent studies on user needs and on more frequent searches for information in lexicographical reference works indicate that users prioritize networked dictionaries and the modular structure of information. These studies also identify the difficulties in actively managing the translation equivalents, as well as their differentiation and delimitation, as the main lack of bilingual resources (Domínguez/Valcárcel 2015, Mann 2010, Meyer/Gurevych 2012, Müller-Spitzer/Koplenig/Töpel 2012).

2.2 A noted valency dictionary

PORTLEX provides detailed information on the nominal phrase from the point of view of valency grammar[2]. Thus, arguments and semantic roles constitute essential work variables in the previous linguistic analysis and, therefore, first-order elements in the entries microstructure. In addition to this, PORTLEX also gives several examples for each possible formal realization of a nominal argument or a combination of arguments.

Examples are manually extracted from different reference corpora for each of the languages involved[3]. Furthermore, colors are used to identify the formal realizations of a particular argument[4]. Besides the examples, notes are often included either attached to a formal realization, a semantic feature, a combination of arguments, or even to a whole lexeme (Figure 1). These notes can refer to a single language or be contrastive, i.e. they can compare results in two or more languages (Figure 2).

2.3 A modular, multilingual and cross-lingual dictionary

The PORTLEX dictionary covers six languages contrasted with each other. Indeed, its database is designed to include more languages. It contains a specific module for each language in which data relating to each one of them is stored. These modules are linked to each other through a mother dictionary (Gouws 2014) where Spanish is the pivot language. This allows the alignment of the data of each language and enables their contrastive display according to the user's needs. In this way, PORTLEX can be defined not only as a multilingual dictionary[5], but above all also as a cross-lingual dictionary (Domínguez 2017: 190–196).

[2] From a typological-lexicographic approach, PORTLEX is a syntagmatic dictionary with special application in linguistic production (Schumacher 2006a, b, Kühn 1989, Model 2010).

[3] The corpora used are CREA for Spanish, DeReKo for German, PAISÀ for Italian, FRANTEXT for French and CORGA for Galician. They are all available online, but some of them are access-controlled which prevents the editing team from linking examples back to their source corpus. Furthermore, the orientation of some corpora sometimes makes it impossible to find examples for certain realizations. In this case, we use Sketch Engine or even the web as a corpus with all the precautions that entails.

[4] PORTLEX manages a list of sixteen semantic roles with each semantic role linked to a specific color. Thus, for example, the argument 'that which performs the action' is marked in green in the subentries and examples (Figures 1 and 2).

[5] To define what a multilingual dictionary is, we could use the proposal made by Pedro Fuertes-Olivera and Henning Bergenholtz in the present volume: "Our above reflections allow us to define multilingual lexicography in the era of the internet as the theory and practice of unified and well-connected monolingual, bilingual and multilingual dictionaries using data from a multilingual database. These dictionaries are information tools that cover words, terms, facts, and/or things in

3 Stages of development

The development of a dictionary with the characteristics mentioned above has taken six years. As will be seen, not only has the funding of public and private entities been crucial throughout this process, but also the growing collaboration among users, editors and administrators within a community of practice. Several phases can be distinguished here:

1. The first phase was launched in 2009 with CSVEA[6], a project coordinated by María José Domínguez and Stefan Schierholz with the aim of preparing the corpus for a Spanish <> German contrastive dictionary on the nominal phrase. During this stage, a theoretical framework of analysis was adopted and a microstructure was established that served as a basis for PORTLEX (Domínguez/Mirazo/Valcárcel forthcoming, Mirazo 2014, 2015).
2. In a second phase (2013–2015) the elaboration of PORTLEX began, coordinated by María José Domínguez and including Galician, Italian and French, as well as Spanish and German. Specialists in these languages joined the team and were trained in the analysis and drafting procedures of the dictionary. In addition, reference dictionaries and corpora were selected and a fundamental aspect of PORTLEX was developed: the structure of its relational database and the visualization of its data in the web interface. During this phase the project was financed by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund 2007–2013 (ERDF)[7]. By the end of this period, and for all the languages involved, 1319 actantial realizations and 1181 combinations had been analyzed, accompanied by some 90 contrastive notes[8].
3. From 2016 onwards, a new phase began marked by the work in areas of application related to the dictionary, while the analysis of nouns continues for the elaboration of new entries. In the didactic field, PORTLEX opened up to the participation of undergraduate or master's students who are training in lexicography or valency grammar, either by consulting the data as basic users or by pre-

several languages, have the same conceptualization at the pre-compilation phase, and make use of lexicographic and technological know-how which allows (a) lexicographers to add new languages to the same information database from which new monolingual, bilingual and multilingual dictionaries can be extracted and (b) users to retrieve connected data easily and to spot and understand possible similarities and differences among the several languages covered".

[6] Spanisch-deutsche Substantivvalenzwörterbuch (CSVEA). The project received financing from the Galician government (XUNTA: INCITE 09 204 074 PR).

[7] The complete reference of the financed project is Portal Lexicográfico (PORTLEX): Diccionario online modular multilingüe y corpus informatizado anotado de la frase nominal (FFI2012-32456) and was coordinated by María José Domínguez (University of Santiago de Compostela).

[8] Notes included in the actant field are not counted here.

paring entries as editors under the supervision of members of the editing team[9]. On the other hand, the dictionary is now also being used by different researchers in linguistics to obtain data for their research. All of these new users, together with the editing team, form a virtual community around PORTLEX where members learn from each other to enter, review or correct data. This stage was consolidated in 2017 with the BBVA Foundation providing financing to develop a prototype of a generator of nominal phrases using the data available in PORTLEX[10]. The project will be carried out by the dictionary editing team and different specialists from its virtual community.

4 The theory behind PORTLEX: the noun valency

Whether from a monolingual, bilingual, or multilingual perspective, the study of the noun and its combinatorial potential still constitutes a line of research in linguistics with important gaps[11]. In our view, this is mainly due to a lack of consensus on the nature of the noun valency. In general, deverbal and deadjectival nouns are commonly recognized as valency carriers, since they often inherit the valence of the words they derive from. However, there are already several studies dealing with the valency of other types of nouns. Generally, works that study the nominal valency address either qualitative features (Alexiadou 2001) or semantic aspects (Ehrich/Rapp 2000), as well as its consideration as a particular system (Teubert 1979: 79, Kubczak/Schumacher 1998: 284–285). However, the combinatorial possibilities of the nominal arguments have not yet received the same attention from researchers (Schumacher 2006: 1442, Iroaie 2008).

PORTLEX was created precisely to study the noun valency, but from a multilingual and lexicographic perspective. The dictionary analyzes not only deverbal (EVALUACIÓN, INVESTIGACIÓN, etc.)[12] and deadjectival nouns (SINCERIDAD, TRANQUILIDAD, etc.), but also non-derivative nouns that present a clear valency nature such as PROBLEMA, GANA or others. The theoretical framework used for the analysis of the

[9] In this regard, collaboration with the Erasmus Mundus Master in Lexicography (EMLex) should be mentioned.

[10] The complete reference of the project is *Generación multilingüe de estructuras argumentales del sustantivo y automatización de extracción de datos sintáctico-semánticos*, financed by the BBVA Foundation Grants for Scientific Research Teams 2017 and also coordinated by María José Domínguez.

[11] Added to this is the lack of consensus over the very valency nature of the noun. Since there are authors who do not recognise it. Nominal valence is not only ignored in the classical works of dependency grammar (Lazard 1994), but some recent authors such as Eisenberg (2001: 262) even state that nouns are not valency carriers in the same sense as verbs.

[12] We refer only to lexemes in Spanish, as it is the pivot language in the “mother” dictionary.

argument structures of all these nouns and their equivalents in the languages involved is that developed by María José Domínguez (2011) in her contrastive study on Spanish and German. On the one hand, a series of roles are defined to identify the semantic function of the nouns' arguments (e.g. 'that which performs an action', 'that which is affected', etc.) as well as their syntactic function (*subiectivus*, *obiectivus*, etc.) and, on the other hand, a list of semantic features ('animate', 'institution', 'object', 'situation', etc.)[13] associated with them and that are present in the different formal realizations of each argument. That is to say, the head of an argument with a certain semantic role can only present specific semantic features, regardless of its formal realization (prepositional phrase, adjective phrase, apposition, compound name, etc.). For example, in the case of the Spanish lexeme TEXTO ('text') there is a complement with the role of 'that which performs the action' with two possible realizations: a prepositional phrase introduced by *de* (*el texto del poeta*) or an adjective (*los textos cervantinos*). In both cases, only the semantic features 'human' and 'institution' can be present and, thus, we understand that the authors of the texts are the poet and Cervantes. However, if one were to say *el texto del cangrejo* ('crab') or *el texto paisajístico* ('landscape'), it would not be understood that a crab or a landscape had written or published a text, but that the text is about them. The reason for this is that CANGREJO presents the semantic feature 'non-human animate' and PAISAJÍSTICO the feature 'immaterial", both associated within the predicate of texto to arguments with the role 'that which is not affected: theme', which also has, among others, adjectival and prepositional phrase realizations. These cases, in which two arguments of a noun with different semantic roles present similar formal realizations, are quite frequent. In fact, on many occasions the only way to know to which argument an actantial realization belongs is to observe its semantic feature.

This approach to noun valence covers two central aspects: on the one hand, it describes the noun and its combinatorial potential, and on the other hand, it provides essential information for the production of new examples, thus overcoming some of the limitations inherent to the examples provided in grammars. That is, without this syntactic-semantic description a user cannot know if other examples that they want to build are possible, they only know that those described in a given resource are. Thus, Kubczak (in Dušek 2013: 131) points out the following in relation to valency dictionaries:

> Sie sind weniger eine Hilfe beim Verstehen von unbekannten Wörtern, als beim Bilden eines korrekten Satzes, wenn sich der Schreiber schon für ein Wort entschieden hat (meistens für ein Verb oder ein Substantiv) und unsicher ist, wie es weiter geht im Satz.

[13] For this purpose, syntactic-formal reformulation tests (e.g. subiectivus vs. obiectivus) are applied, as well as anaphor and question tests. The complete lists of semantic roles and features used is provided in the dictionary user manual (PORTLEX 2015) and they are based on the proposals made by Engel (2004) and Domínguez (2011).

In a laborious process, PORTLEX gathers all the possible realizations of a noun argument and documents them with examples extracted from corpus and websites. Thus, in addition to the so-called *complemento del nombre* (CN) in the Spanish-speaking grammatical tradition, a prepositional phrase normally introduced by *de* in Spanish (i.e. *el libro del periodista*), genitive pronouns (Rigau 1999: 340) (*el tema del libro* = *su tema*; 'the topic of the book' = *its* topic) are also possible[14], as well as adjectives (i.e. *la actividad fotosintética* in Spanish), compounds (i.e. *Trennungsratschlag* in German), N_1N_2 realizations (i.e. *sabor chocolate* in Spanish or *tendance musique* in French) (RAE 2010: 192, Valcárcel 2017), and different kinds of sentence realizations (i.e. *el deseo de viajar* in Spanish or *una discussione su come fare il caffè* in Italian). This means that a noun complement can be realized at word level –in some languages as a compound– at phrasal level or at sentence level[15].

In addition to its formal realizations, each nominal argument also presents a syntactic function (Rigau 1999: 340–341). Thus, in the Spanish phrase *el texto del poeta* the complement plays a function of subject, while in *el consumo energético* the adjectival realization of the complement presents a function of object. It is distinguished, therefore, between subject complement, object complement, prepositional complement, adverbial complement, expansive complement, and nominal complement. In this way, PORTLEX describes the noun valency at all its levels, from semantics to superficial realizations, passing through its deep syntactic actants, according to Igor Mel'čuk (2015: 63–66) terminology, that is, the syntactic functions.

5 Innovations

As with other valency or multilingual dictionaries, PORTLEX has been conceived to help in different linguistic mediation processes (Mirazo 2016: 94–95). Moreover, this dictionary also attempts to respond to different needs relating to the processing of multilingual linguistic information on the valency of the noun phrase, the cross-lingual visualization of this information and a more active role of dictionary users in updating the dictionary. In this respect, the innovations brought by PORTLEX could be summarized in the following aspects.

[14] In the case of German, it is also important to take into account the frequent use of the genitive case (i.e. die Krankheit der Schafe, 'the disease of sheep').

[15] For sentence level complements see Domínguez Vázquez/Mirazo Balsa (2017).

5.1 Comprehensible information on the argument structure of the noun phrase

Considering that, since Tesnière, dependency grammar has always been strongly anchored on a verb-centered approach, it does not seem strange that most of the monolingual and bilingual valency dictionaries available focus on verb valency. In spite of this, there are different dictionaries on the valency of the noun phrase available for several of the languages with which PORTLEX is concerned. To date, only one single monolingual dictionary is available for the study of the noun valency, the *Wörterbuch zur Valenz und Distribution der Substantive* (Sommerfeldt/Schreiber 1977), which constitutes an essential reference for the German language. Among the bilingual dictionaries we should also mention those of Bassola et al. (2003, 2009) and Angelini/Fábián (2005). Nor must we forget to mention some general valency dictionaries considering the argument structure of nouns such as the DEC (Mel'čuk 1982–1999), popularized by the *Lexique actif du français* (Mel'čuk/Polguère 2007), or the DCF for French[16]. For the Spanish language, the DiCE also presents detailed data on the argument structure of nouns, although its aim is to provide collocational information.

PORTLEX joins these and other works to overcome the lack of lexicographic tools focused on the noun valency. However, it stands out amongst valency dictionaries for a series of features already mentioned here: it is multilingual and cross-lingual, it is available online and, above all, it is not a finished work, but is constantly updated thanks to its collaborative nature. From the strictly valential point of view, PORTLEX distinguishes itself by using didacticized semantic roles (e.g. 'that which performs the action' instead of 'agent'). As already mentioned, each semantic role is also marked with a specific color. In this sense, PORTLEX follows the guidelines adopted for VALBU or FrameNet (Ruppenhofer et al. 2016) since, among other reasons, it considerably eases the search and comprehension of noun arguments by the user. On the other hand, and following the work of Bassola et al. (2003, 2009), PORTLEX facilitates the different realizations for each actant or argument in the five languages covered, avoiding the display of abstract schemes of the noun valency. Finally, PORTLEX entries present an exhaustive section on the combinations of all arguments among themselves.

[16] Learner's dictionaries containing information on the noun's argument structure are becoming increasingly frequent. As an example, we might mention the DAFLES for French.

5.2 A multi/cross-lingual dictionary

As previously mentioned, PORTLEX is a multilingual dictionary that covers five languages. However, the fact that it is possible to access the argument structure of a noun in one language and contrast it on screen with that of its equivalents in other languages makes it a real cross-lingual tool. This has been made possible mainly thanks to the modular structure of the relational database that feeds PORTLEX. Thus, each language has a module in the database and the connection between them is established through a common meaning, which serves as an index. In this respect, PORTLEX operates in a similar way to other multilingual tools[17], though in the field of valency lexicography and while substantially improving the visualization of data: it is the user who decides at all times which data and which languages they wish to compare. Beyond the importance of other multilingual tools, however, the work of Rufus Gouws and his concept of "mother dictionary" (Gouws 2014) played a key role in all these developments.

In the case of PORTLEX, the mother dictionary was constituted by a list of meanings extracted for the lexemes analyzed in the Spanish language, which also provides the metalanguage of the dictionary, during the first phase of the project (section 2). These meanings or definitions form the keys of interlinguistic relation within the database and also serve to select the equivalents to be analyzed in the different languages. Among other criteria, in order to justify the selection of an equivalent in a given language, that equivalent must have an equivalent meaning in the respective reference dictionaries of that language[18]. Obviously, this methodology is not exempt from problems, as pointed out by Mirazo (2016: 99–101), which are mainly related to the lack of full equivalence between all languages[19]. However, beyond its limitations, operating with the concept of mother dictionary and adopting meaning as a relational key have made possible the versatile cross-lingual visualization offered by PORTLEX today (section 7.3).

[17] Wordnet-based multilingual tools, for example, use the so-called "synsets", sets of cognitive synonyms grouped by a common definition or gloss, as indexes to link equivalencies in different languages. The synsets were established for English in the original Wordnet developed at Princeton University (Solla/Gómez 2015: 171–172).

[18] The reference dictionaries used in PORTLEX are CLAVE and DLE for Spanish, DWDS and DUDEN for German, GARZANTI and TRECCANI for Italian, TLFI and LPR for French and DRAG for Galician. Definitions taken from reference dictionaries are shown in italics, while those provided by the PORTLEX editorial team appear in regular text. In some instances, both the dictionary definition and the PORTLEX one are given to aid comprehension and the selecting of entries by the user.

[19] However, since it is based on Spanish, PORTLEX presents the data of the other languages as equivalent to Spanish. This should be especially taken into account when carrying out contrastive queries between languages other than Spanish (Mirazo 2016: 100).

5.3 Collaborative updating

As already noted above, PORTLEX is a semi-collaborative dictionary as users can update it although always under the supervision of the editing team (Abel/Meyer 2013, Valcárcel forthcoming). PORTLEX is therefore not a closed work, but is regularly updated and expanded within the framework of a community of users sharing common interests and goals. As in any other virtual community, it is possible to distinguish several types of members within PORTLEX (Kim 2000). Since users must register to have free access to the dictionary, a first distinction should be made between registered and non-registered members. Non-registered members are basically followers of our website[20] and our social media profiles. However, they are extremely important for maintaining the online reputation and digital identity of the project, as well as to attract valuable new users to the dictionary.

There are currently 80 registered users on PORTLEX. All these users constitute the core of a virtual community comprising specialists in different languages and fields of linguistics, masters and doctoral students, as well as all the members of the editing team. Registered members can be divided into administrators, reviewers, and users according to their roles when using PORTLEX. Users mainly consult the dictionary but they can also take part in the PORTLEX discussion group along with the editors, the reviewers and even with the administrators. Editors are trained researchers, specialists in a particular language, who edit entries in the dictionary database, which are in turn checked by the reviewers. All editors and reviewers received specific training, not only on the functioning of the database, but also on valency grammar. Finally, administrators can edit and delete content as well, but they can also create, edit, or even delete user profiles if needed.

This social component is not really an innovation in lexicography. In fact, it is directly inspired by the wiktionaries and WordReference (Meyer/Gurevych 2012, Valcárcel forthcoming). But unlike these projects, the goal of PORTLEX is not so much the creation and maintenance of a large community of users, but to enlarge the community of experts on valency grammar and lexicography involving them in the development of a multilingual dictionary. That implies that users, if they want to, can become editors of the dictionary after a training process. This has already been the case for several students who, for their academic projects, have drafted PORTLEX entries in different languages. Some of them have even entered their data directly into the database and had their work properly certified by the administration team. In this way, PORTLEX not only proves to be a useful tool for linguistic

[20] [<https://diccionarioportlex.wordpress.com>].

consultation or research, but also for the practical learning of valential grammar and, in general, lexicographic techniques[21].

6 A relational database for a modular structure

The specific features of PORTLEX described above not only implied a deep reflection on the organization of the data in its entries, but above all a reformulation of its basic structure and, inevitably, the development of a custom-made database. This section will address these three issues to explain how the design of the database had to be customized to both the structure and microstructure of the web interface of the dictionary.

6.1 A modular structure

To enable PORTLEX to operate as a true multi/cross-lingual dictionary, a modular structure was adopted to ensure that the data related to each language would be recorded and presented separately. Each language has its own module, a specific dictionary where the semantic roles of the arguments of every lexeme, as well as their possible combinations, are analyzed in detail. All this information about a lexeme in one language module is linked and aligned to that of its equivalents in the rest of the modules. Those links between equivalent lexemes are made by means of the above-mentioned mother dictionary that contains the common signified they share[22]. Besides cross-lingual visualization, this modular structure allows monolingual queries and, in fact, PORTLEX provides different ways to visualize the existing content for a given lexeme in a single module or language (section 7).

[21] This positive experience has led the project team to think about consolidating the PORTLEX dimension as a learning tool. It is expected that an online course will be implemented on a learning platform to not only better train users who wish to edit and enter data into the dictionary, but also learn more of lexicography and valency grammar. At the end of their training, these user-learners would have to submit a final project consisting of the drafting of a dictionary entry. This training would be certified and the drafting could be included in the database. A first step in this direction has already been taken with an academic project presented by Olha Novikova for the EMLex Master. She developed a series of video tutorials and learning activities aimed at facilitating the consultation and understanding of the data provided by PORTLEX.

[22] For more details about the mother dictionary and the cross-lingual functioning within PORTLEX, see 5.2 and 7.3.

6.2 An understandable microstructure

In a monolingual search, the entry that PORTLEX displays for each lexeme basically comprises four main areas or zones:

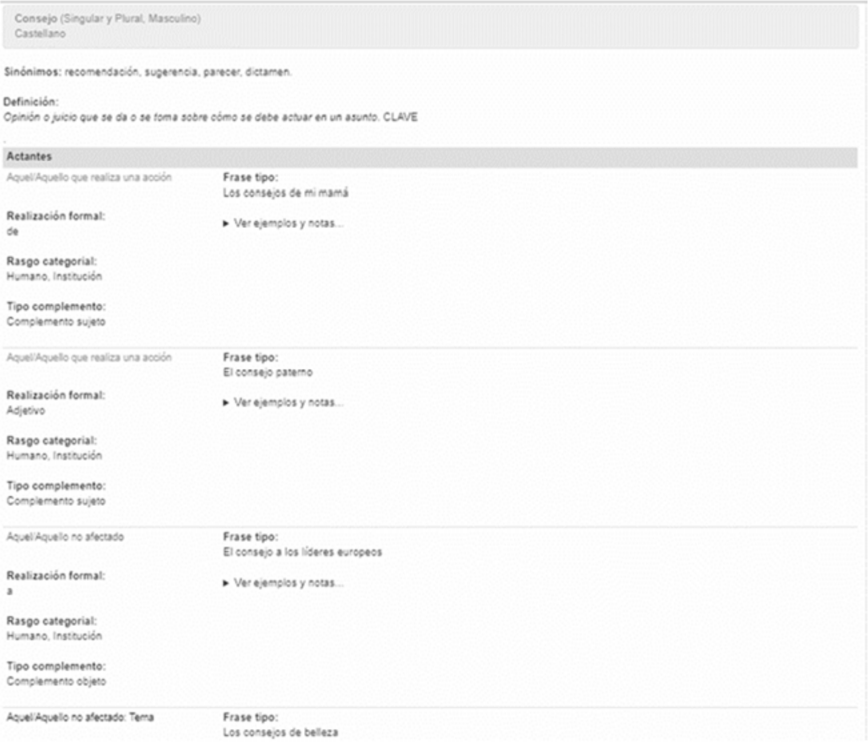
a. Semantic/morphological zone. Placed at the top of the entry, it focuses on the basic semantic and morphological definition of the lexeme. Apart from a list of synonyms[23], this zone contains the definition of the equivalent in one or more reference dictionaries. Morphological information is limited to gender and number.

b. Semantic/syntactic zone. Named Actantes in the entry, this zone describes the elements (actants or arguments) that the lexeme can semantically and syntactically govern. As already explained (sections 4 and 5.1), the semantic role and the different formal realizations for each actant are provided, indicating for each of them the syntactic structure (preposition + noun, noun, adjective, etc.), the syntactic function (subject complement, nominal complement, prepositional complement, etc.), and the semantic feature of its head. This section includes an illustrative sub-zone showing a sample phrase and different examples extracted from reference corpora.

c. Co-occurrence zone (Combinaciones). In this section the different combinations of arguments that have been found for a lexeme in the reference corpora are provided. For each combination, the formal realizations of each involved argument are indicated in order of co-occurrence, identifying for each of them the dominant semantic feature and the syntactic function. This section also includes an illustrative sub-zone with sample phrases and examples of corpora.

d. The additional information zone (Otras clasificaciones). This zone provides different types of information considered relevant by the editing team. This often includes information of a contrastive nature, i.e. on the specificities of a lexeme in one language with respect to its equivalents in the other languages of the dictionary. Usually such information relates to syntactic-semantic restrictions, but in some cases additional information of a lexical, stylistic or pragmatic nature is also presented. There is also information on collocations and phraseology in this zone.

[23] Synonyms can also be accessed in a specific module in PORTLEX.



Consejo (Singular y Plural, Masculino)
Castellano

Sinónimos: recomendación, sugerencia, parecer, dictamen.

Definición:
Opinión o juicio que se da o se toma sobre cómo se debe actuar en un asunto. CLAVE

.

Actantes

Aquel/Aquello que realiza una acción

Frase tipo:
Los consejos de mi mamá

Realización formal:
de

▸ Ver ejemplos y notas...

Rasgo categorial:
Humano, Institución

Tipo complemento:
Complemento sujeto

Aquel/Aquello que realiza una acción

Frase tipo:
El consejo paterno

Realización formal:
Adjetivo

▸ Ver ejemplos y notas...

Rasgo categorial:
Humano, Institución

Tipo complemento:
Complemento sujeto

Aquel/Aquello no afectado

Frase tipo:
El consejo a los líderes europeos

Realización formal:
a

▸ Ver ejemplos y notas...

Rasgo categorial:
Humano, Institución

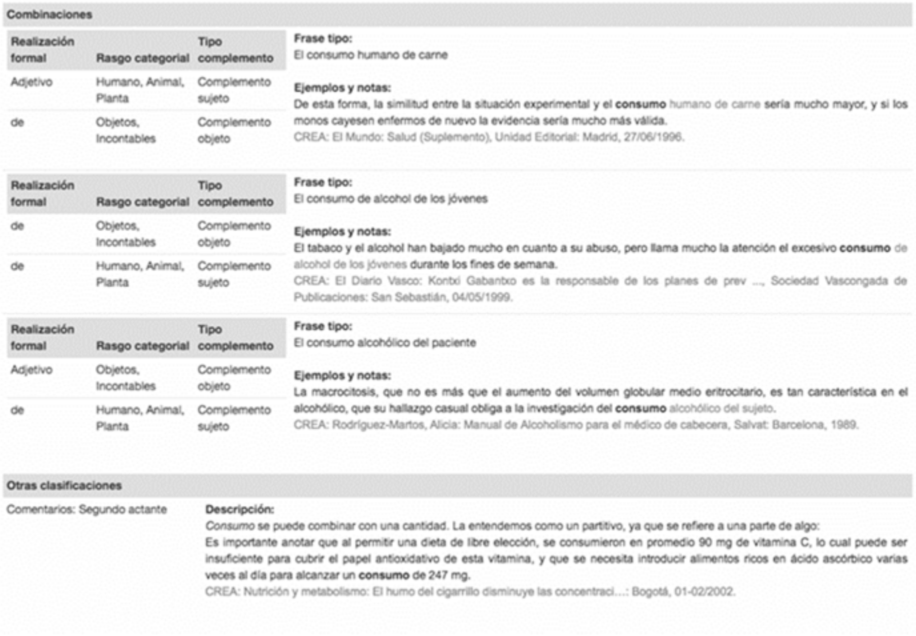
Tipo complemento:
Complemento objeto

Aquel/Aquello no afectado: Tema

Frase tipo:
Los consejos de belleza

Fig. 1: View of the semantic/morphological and semantic/syntactic zones for the Spanish lexeme consejo1 ('advice')

As can be seen in Figures 1 and 2, the PORTLEX microstructure follows several strategies to didactize the presentation of information and to facilitate its global and analytical comprehension by users. Besides what has already been mentioned concerning the use of semantic roles and their chromatic marking, especially useful to highlight arguments in the examples, other graphic (use of tables, lines and bars to outline the different sections and subsections of the entry), typographical (abbreviations are avoided) and illustrative procedures (inclusion of sample phrases and examples from corpora) should also be mentioned.



Combinaciones

Realización formal	Rasgo categorial	Tipo complemento
Adjetivo	Humano, Animal, Planta	Complemento sujeto
de	Objetos, Incontables	Complemento objeto

Frase tipo:
El consumo humano de carne

Ejemplos y notas:
De esta forma, la similitud entre la situación experimental y el **consumo** humano de carne sería mucho mayor, y si los monos cayesen enfermos de nuevo la evidencia sería mucho más válida.
CREA: El Mundo: Salud (Suplemento), Unidad Editorial: Madrid, 27/06/1996.

Realización formal	Rasgo categorial	Tipo complemento
de	Objetos, Incontables	Complemento objeto
de	Humano, Animal, Planta	Complemento sujeto

Frase tipo:
El consumo de alcohol de los jóvenes

Ejemplos y notas:
El tabaco y el alcohol han bajado mucho en cuanto a su abuso, pero llama mucho la atención el excesivo **consumo** de alcohol de los jóvenes durante los fines de semana.
CREA: El Diario Vasco: Kontxi Gabantxo es la responsable de los planes de prev ..., Sociedad Vascongada de Publicaciones: San Sebastián, 04/05/1999.

Realización formal	Rasgo categorial	Tipo complemento
Adjetivo	Objetos, Incontables	Complemento objeto
de	Humano, Animal, Planta	Complemento sujeto

Frase tipo:
El consumo alcohólico del paciente

Ejemplos y notas:
La macrocitosis, que no es más que el aumento del volumen globular medio eritrocitario, es tan característica en el alcohólico, que su hallazgo casual obliga a la investigación del **consumo** alcohólico del sujeto.
CREA: Rodríguez-Martos, Alicia: Manual de Alcoholismo para el médico de cabecera, Salvat: Barcelona, 1989.

Otras clasificaciones

Comentarios: Segundo actante

Descripción:
Consumo se puede combinar con una cantidad. La entendemos como un partitivo, ya que se refiere a una parte de algo:
Es importante anotar que al permitir una dieta de libre elección, se consumieron en promedio 90 mg de vitamina C, lo cual puede ser insuficiente para cubrir el papel antioxidativo de esta vitamina, y que se necesita introducir alimentos ricos en ácido ascórbico varias veces al día para alcanzar un **consumo** de 247 mg.
CREA: Nutrición y metabolismo: El humo del cigarrillo disminuye las concentraci...: Bogotá, 01-02/2002.

Fig. 2: View of the co-occurrence and additional information zones for the Spanish lexeme CONSUMO^1 ('consumption')

6.3 A custom-made database

The implementation of PORTLEX involved the development of a multilingual database from which a cross-lingual valency dictionary with a very specific microstructure for entries could be generated. This entailed including in the working team programmers with experience, on the one hand, in the development of relational databases and, on the other hand, in lexicographic projects. Finally, Alberto Simões and Mário Vale, from the IPCA (Instituto Politécnico do Cávado e do Ave, Portugal), assumed the task of constructing a custom-made database for PORTLEX. Given the technical specificity of the project, the development of the database involved a short training process for programmers in basic notions of valency grammar and multilingual lexicography. Since 2015, the PORTLEX database has been hosted on a private server and is freely available to all registered users. At a technical level, it runs on Linux and it was developed in PHP using MySQL as the management system and Yii as the web application framework. Through the web interface that makes up the dictionary, the user consults the database using a series of standard MySQL queries (section 7 for more information). However, administrators can perform more complex queries on demand from other users.

The database is updated periodically throughout the collaborative work process explained above (section 5.3). The workflow for inputting data into the database is initiated by entering a lemma or headword representing a definition of the mother dictionary. This lemma or headword, after validation by the database administrators, is the relational key that will connect the different data related to this meaning in the different working languages. From there, the introduction of data for each language is done progressively, going through different revision and validation stages that concern the choice of the right equivalent, its synonyms and its justification with a reference definition, as well as the introduction of data for each argument realization and combination.

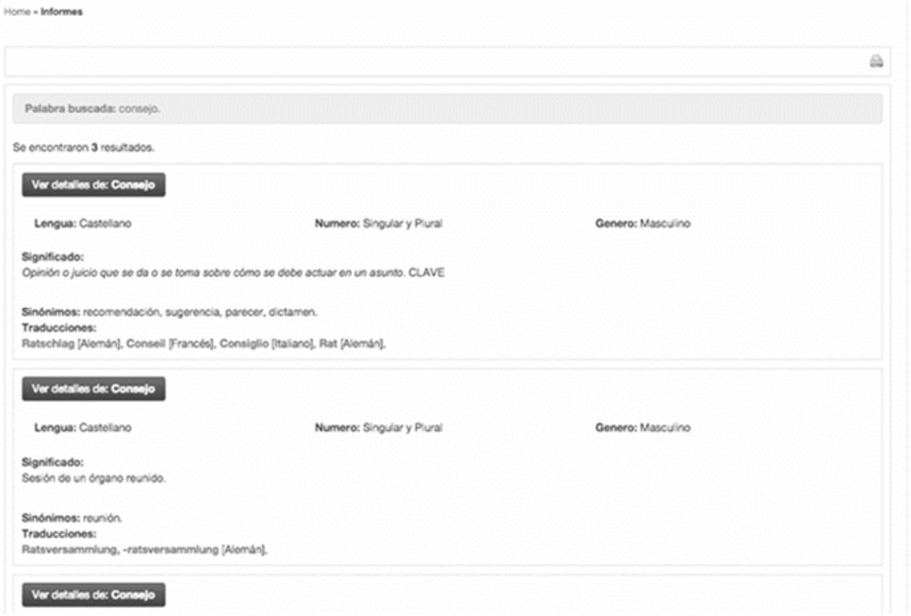
In order to guarantee the correct inter-linguistic alignment of argument realizations or combinations in cross-lingual consultations, common numerical codes are used for each typology. Thus, for example, the realizations [*de* | *di* | *von*] + nominal phrase of the argument 'that which performs the action' share a common code. This allows not only to visually align the equivalent data in two languages, but also to leave empty slots in any language that presents no equivalence of any kind. This is the case, for example, of the French N_1N_2 realizations (*consommation vin*, *une maladie poumon*), which rarely have a similar equivalent in Spanish (Valcárcel 2017). In the cross-lingual display, an empty gap appears in this language next to the French N_1N_2 realizations. Furthermore, these alignment codes also allow the information to be presented respecting the order of combination, which means, the information about the argument that is described first in the semantic/syntactic zone will appear first in the co-occurrence zone.

7 Searches and visualization: retrieving information

At any time, the user of PORTLEX can decide the level of detail and multilingual contrastivity of his or her search. In any case, the user will have to search first for a lexeme or headword. Users will then receive simple monolingual information. From this stage, the user must decide if he or she wants more detailed information about the actants, their different realizations and possible combinations, and if users want to see the available examples. Moreover, users can compare, in one single window, all this information with that of another language of the dictionary.

7.1 Simple search

The user can search directly for nouns in Spanish, Galician, German, French, or Italian. In most cases, the dictionary will give the user a list of different entries for the searched noun. For each entry or noun, the user will find its language[24] and all the information contained in the semantic/morphological area (section 6.2), as well as the equivalents in the languages involved in the dictionary. Clicking on each of these equivalents will open its entry. In some cases, more than one equivalent can be shown for a language. This type of query is especially useful for those users who are proficient in only one of the languages of the dictionary but who want to consult data on the equivalents in other languages.



Palabra buscada: consejo.

Se encontraron 3 resultados.

Lengua: Castellano Numero: Singular y Plural Genero: Masculino

Significado:
Opinión o juicio que se da o se toma sobre cómo se debe actuar en un asunto. CLAVE

Sinónimos: recomendación, sugerencia, parecer, dictamen.
Traducciones:
Ratschlag [Alemán], Conseil [Francés], Consiglio [Italiano], Rat [Alemán].

Lengua: Castellano Numero: Singular y Plural Genero: Masculino

Significado:
Sesión de un órgano reunido.

Sinónimos: reunión.
Traducciones:
Ratsversammlung, -ratsversammlung [Alemán].

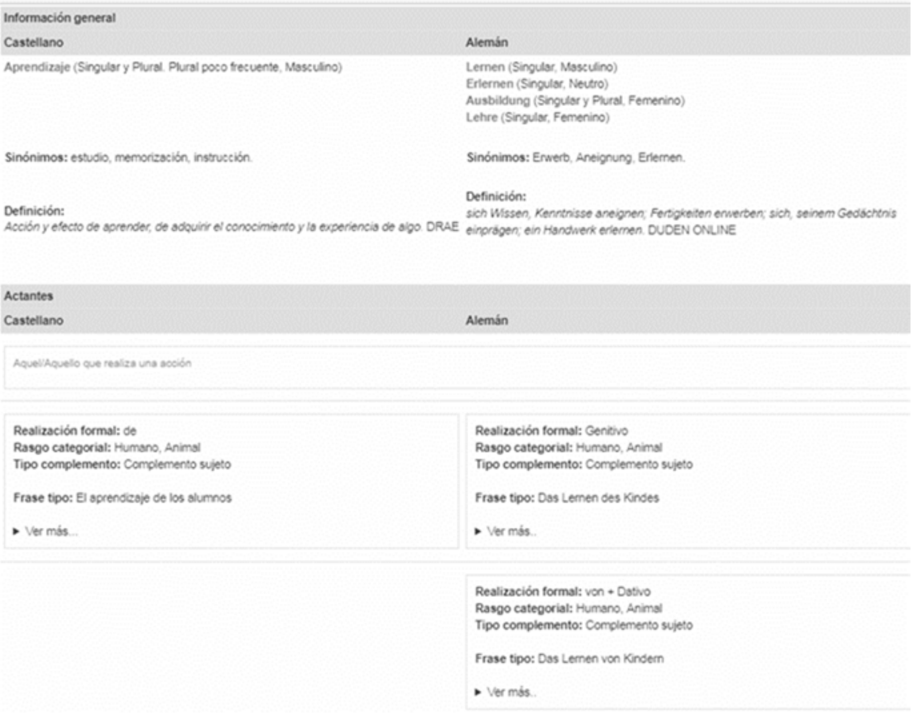
Fig. 3: Partial view of the simple search results for the wordform consejo ('advice')

7.2 Detailed search

Simple searches also give access to more detailed information about a lexeme. For each listed entry there is a *Ver detalles* ('See Details') button to access detailed in-

[24] There may be entries whose headword is a homograph in different languages. For example, odio ('hate') has a common spelling in Spanish and Galician, the same as morte ('death') for Italian and Galician.

formation about the complements and combinations for that entry. For each detailed search, PORTLEX displays the information for all the microstructure areas described in section 6.2. To avoid information overload in this type of detailed view, examples and notes are displayed only when the user clicks on *Ver ejemplos y notas*, an option available for each argument realization and combination (Figures 1 and 2).



Información general

Castellano	**Alemán**
Aprendizaje (Singular y Plural. Plural poco frecuente, Masculino)	Lernen (Singular, Masculino) Erlernen (Singular, Neutro) Ausbildung (Singular y Plural, Femenino) Lehre (Singular, Femenino)
Sinónimos: estudio, memorización, instrucción.	**Sinónimos:** Erwerb, Aneignung, Erlernen.
Definición: *Acción y efecto de aprender, de adquirir el conocimiento y la experiencia de algo.* DRAE	**Definición:** *sich Wissen, Kenntnisse aneignen; Fertigkeiten erwerben; sich, seinem Gedächtnis einprägen; ein Handwerk erlernen.* DUDEN ONLINE

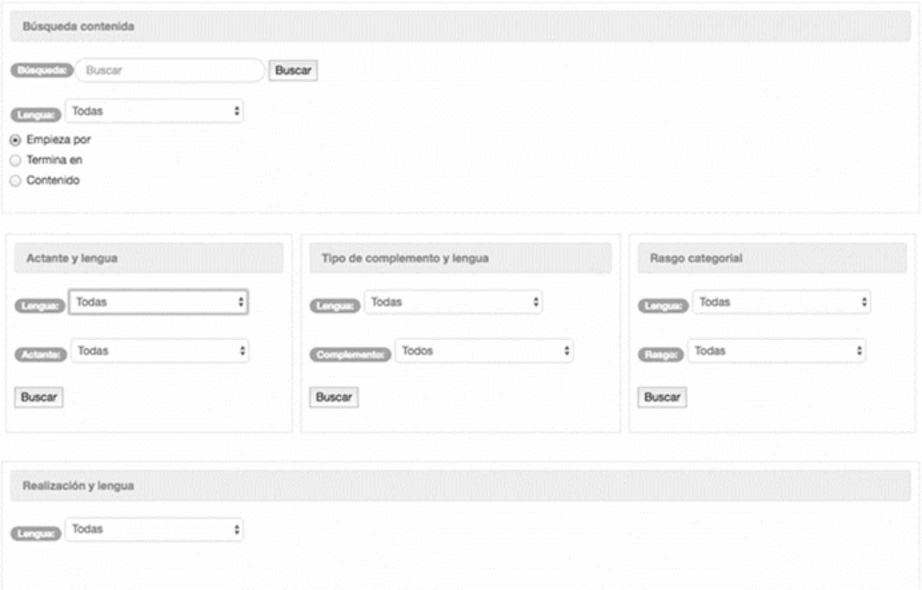
Actantes

Castellano	**Alemán**
Aquel/Aquello que realiza una acción	
Realización formal: de **Rasgo categorial:** Humano, Animal **Tipo complemento:** Complemento sujeto **Frase tipo:** El aprendizaje de los alumnos ▸ Ver más...	**Realización formal:** Genitivo **Rasgo categorial:** Humano, Animal **Tipo complemento:** Complemento sujeto **Frase tipo:** Das Lernen des Kindes ▸ Ver más..
	Realización formal: von + Dativo **Rasgo categorial:** Humano, Animal **Tipo complemento:** Complemento sujeto **Frase tipo:** Das Lernen von Kindern ▸ Ver más..

Fig. 4: Comparative map for the Spanish lexeme aprendizaje1 (‘learning’) with lernen, one of its German equivalents. Partial view of the information on actants

7.3 Cross-lingual display

Users can compare detailed information about a noun in a language with that of its equivalent in another language. In order to do this, they must first search for the noun in one language and open the detailed information. At the beginning and end of the results table they will find the *Vista contrastiva* (comparative display) section (Figure 4). To see the comparative map in another language, the users must click on the corresponding language link. The user will then access a comparative table, which shows the information about the searched lexeme in the right-hand column and the information about its equivalent in another language in the left-hand column. The format and type of information given about arguments and their combinations is similar to that obtained from the detailed search. Additionally, links to other possible equivalents for that entry are also offered. All the information is shown in rows, so that if an argument or a combination of complements does not exist in a language, there will be a blank space in the row corresponding to that language. However, in addition to formal equivalence, information is also given concerning semantic parallelism. The user can always return to the monolingual view by clicking on the corresponding noun link.



Búsqueda contenida

Búsqueda: Buscar | Buscar

Lengua: Todas

● Empieza por
○ Termina en
○ Contenido

Actante y lengua

Lengua: Todas

Actante: Todas

Buscar

Tipo de complemento y lengua

Lengua: Todas

Complemento: Todos

Buscar

Rasgo categorial

Lengua: Todas

Rasgo: Todas

Buscar

Realización y lengua

Lengua: Todas

Fig. 5: Partial view of the advanced search screen with its different sections

7.4 Advanced search

In addition to the possibilities mentioned above, users can carry out more specific searches by clicking on *Búsqueda avanzada* ('Advanced Search'). This opens a page with different search fields where it is possible to retrieve the list of all analyzed lexemes in the dictionary (in one single language or in all the languages involved) and, above all, the lists of lexemes with a specific argument, complement type, semantic feature, or formal realization in one of the dictionary's languages. As mentioned before, it is also possible to obtain results for more complex queries on demand.

8 Conclusion

The development of a multilingual dictionary in line with the features of PORTLEX (cross-lingual, modular and valency oriented) represented a major challenge that required the effort of a multidisciplinary team. Specialists in valency grammar of different languages and from different linguistic traditions, experts in lexicography, and system programmers have worked side by side, learning from each other, to carry out this project. The result is a very flexible tool that allows, on the one hand, complex queries to retrieve very precise contrastive data on the noun phrase in five languages and, on the other hand, the building of a community of experts in valency grammar based on their work with PORTLEX. In fact, the updating and expansion of the database is increasingly fuelled by the involvement of the user community. Several students and young researchers have already collaborated on the analysis of lexemes, data entries, or the development of learning materials that help to better understand the potential of PORTLEX and the data it offers. Precisely in relation to the latter, and beyond the need for developing specific training resources, there are now two new challenges ahead: the optimization of PORTLEX by combining it with other available tools and the application of the data it provides in the development of new resources. The MultiGenera project, for which the PORTLEX team has received funding from the BBVA Foundation, has just been launched to respond to both these challenges.

9 Bibliography

9.1 Dictionaries

Angelini/Fábián = Angelini, Maria Teresa/Fábián, Zsuzsanna, *Dizionario italiano-ungherese della valenza dei nomi*. Szeged: Grimm Kiadó, 2005.

Bassola et al = Bassola, Peter et al., *Deutsch-ungarisches Wörterbuch zur Substantivvalenz*. Szeged: Grimm Kiadó, 2003, 2009. 2003=Volume 1, 2009= Volume 2.

CLAVE = *Diccionario CLAVE. Diccionario de uso del español actual*. Ed. SM. [http://clave.smdiccionarios.com/].

DAFLES = *DAFLES (Dictionnaire d'apprentissage du français langue étrangère et seconde)*. [https://ilt.kuleuven.be/inlato/].

DCF = *Dictionnaire combinatoire du français. Expressions, locutions et constructions*. Zinglé Henri/ Brobeck-Zinglé, Marie-Louise. Paris: La maison du dictionnaire, 2003.

DEC = *Dictionnaire explicatif et combinatoire du français contemporain: recherches lexico-sémantiques I-IV*. Ed. by Mel'čuk, Igor' et al.. Montréal: Presses de l'Université de Montréal, 1984-1999.

DiCE = *Diccionario de Colocaciones del Español*. Ed. Alonso, Margarita. [<http://www.dicesp.com>; last access: October 10, 2017].

DLE = *Diccionario de la lengua española*. Ed. Real Academia Española. [<http://dle.rae.es/>; last access: October 10, 2017].

DUDEN = *Duden online Wörterbuch*. Ed. Duden. [<https://www.duden.de/>; last access: October 10, 2017].

DRAG = *Dicionario da real Academia Galega*. Dir. González, Manuel. Real Academia galega [<http://academia.gal/dicionario>; last access: October 10, 2017].

DWDS = *Digitales Wörterbuch der deutschen Sprache*. Ed. Berlin-Brandenburgische Akademie der Wissenschaften [<https://www.dwds.de>; last access: October 10, 2017].

GARZANTI = *Il grande dizionario Garzanti della lingua italiana*. Ed. Garzanti. [<https://www.garzantilinguistica.it/ricerca/?q=dizionario>; last access: October 10, 2017].

LPR = *Le Petit Robert. Dictionnaire alphabétique et analogique de la langue française.* Dir. Rey, Alain/Rey-Debove, Josette. Paris: Le Robert, 2012.

PORTLEX = *Diccionario multilingüe de la valencia del nombre*. Dir. Domínguez Vázquez, Mª José [<http://portlex.es>; last access: October 10, 2017].

TLFI = *Trésor de la langue française informatisée*. Ed. ATILF (CNRS & Université de Lorraine) [<http://www.atilf.fr/tlfi>; last access: October 10, 2017].

TRECCANI = *Treccani 2014. Dizionario della lingua italiana*. Ed. Istituto della Enciclopedia italiana / Giunti T.V.P. Editori. Roma, 2014.

VALBU = *Valenzwörterbuch deutscher Verben*. Schumacher, Helmut/Kubczak, Jacqueline/Schmidt, Renate/de Ruiter, Vera. Tübingen: Narr, 2004.[25]

9.2 Monographes and articles

Alexiadou, Artemis (2001): *Functional Structure in Nominals*. Amsterdam, Philadelphia: Benjamin's.

Abel, Andrea/Meyer, Christian M. (2013): The Dynamics Outside the Paper: user contributions to online dictionaries. In: Kosem, Iztok/Kallas, Jelena/Gantar, Polona/Krek, Simon/Langemets, Margit/Tuulik, Maria (coords.): *Electronic lexicography in the 21st century: thinking outside the paper: proceedings of the eLex 2013 conference, 17-19 October 2013, Tallinn, Estonia*. Ljubljana, Tallinn: Trojina, Institute for Applied Slovene Studies / Eesti Keele Instituut, 179–194.

Condette, Marie-Hélène/Marin, Rafael/Merlo, Aurélie (2012): La structure argumentale des noms déverbaux : du corpus au lexique et du lexique au corpus. In: *SHS Web of Conferences* 1, 845–58.

De Schryver, Gilles-Maurice (2003): Lexicographers' Dreams in the Electronic Dictionary Age. In: *International Journal of Lexicography* 16 (2), 143–199.

Domínguez, Mª José (2017): Portales y diccionarios multilingües electrónicos. In: Domínguez Vázquez, Mª José/Sanmarco Bande, Mª Teresa (ed.): *Lexicografía y didáctica*. Frankfurt: Peter Lang, 177–201.

Domínguez Vázquez, Mª José (2014): Nomenergänzungen aus grammatischer Sicht: Forschungsstand und Bestandsaufnahme. In: *Neuphilologische Mitteilungen: bulletin de la Société Néophilologique de Helsinki* (1), 3–32.

Domínguez Vázquez, Mª José (2011): *Kontrastive Grammatik und Lexikographie: spanisch-deutsches Wörterbuch zur Valenz des Nomens*. Munich: Iudicium.

[25] The online version (e-VALBU) is available on <http://hypermedia.ids-mannheim.de/evalbu/index.html>.

Domínguez, Mª José/Mirazo, Mónica/Valcárcel, Carlos (forthcoming): Evolución del diccionario bilingüe al multilingüe: de CSVEA a PORTLEX. In: Meliss, Meike/Sánchez, Mª Dolores/Sanmarco, Mª Teresa (ed.): *A lexicografía das linguas románicas: Estado da cuestión*. Munich: Iudicium.

Domínguez Vázquez, Mª José/Mirazo Balsa, Mónica (2017): Aproximación multilingüe a los argumentos oracionales del sustantivo: dificultades y retos. In: Domínguez Vázquez, Mª José/Kutscher, Silvia (ed.): *Estudios contrastivos y multicontrastivos: Interacción entre gramática, didáctica y lexicografía*. Berlin: De Gruyter, 353–367.

Domínguez Vázquez, Mª José/Valcárcel Riveiro, Carlos (2015): 'Hábitos de uso de los diccionarios entre los estudiantes universitarios europeos: ¿nuevas tendencias?'. In: Domínguez Vázquez, Mª José/Gómez Guinovart, Xavier/Valcárcel Riveiro, Carlos (eds.): *Lexicografía de las lenguas románicas II. Aproximaciones a la lexicografía contemporánea y contrastiva*. Berlin: De Gruyter, 165–189.

Dušek, Ondřej (2013): *Zum Vergleich der tschechischen und deutschen Valenzwörterbücher*. [<http://ufal.mff.cuni.cz/~odusek/theses/ma_thesis.pdf>; last access: April 19, 2016]

Ehrich, Veronika/Rapp, Irene (2000): Sortale Bedeutung und Argumentstruktur: -ung- Nominalisierungen im Deutschen. In: *Zeitschrift für Sprachwissenschaft* 19 (2), 245–303.

Eisenberg, Peter (2001): *Grundriss der deutschen Grammatik: Der Satz*. Stuttgart: Metzler.

Engel, Ulrich (2004): *Deutsche Grammatik – Neubearbeitung*. München: Iudicium.

Engel, Ulrich (2009): *Syntax der deutschen Gegenwartssprache.* Berlin: Erich Schmidt Verlag.

Fuertes-Olivera, Pedro/Bergenholtz, Henning (2018): Towards a New Definition of Multilingual Lexicography in the Era of the Internet. In: Domínguez, Mª José/Mirazo, Mónica/Valcárcel, Carlos (ed.): *Studies on Multilingual Lexicography*. Berlin: De Gruyter, 9-29.

Gouws, Rufus (2014): Towards bilingual dictionaries with Afrikaans and German as language pair. In: Domínguez Vázquez, Mª José/Mollica, Fabio/Nied Curcio, Martina (ed.): *Zweisprachige Lexicographie zwischen Translation und Didaktik.* Berlin: De Gruyter, 249–262.

Iroaie, Ana (2008): *Lexikographische Grundlagen zur Beschreibung der Valenz deutscher und rumänischer Substantive*. Universität Bukarest: Bukarest.

Kim, Amy Jo (2000): *Community building on the web: Secret strategies for successful online communities*. Boston: Addison-Wesley Longman Publishing Co., Inc..

Klosa, Annette (2013): The lexicographical process (with special focus on online dictionaries). In: Gouws, Rufus H./Heid, Ulrich/Schweickard, Wolfgang/Wiegand, Herbert Ernst (ed.): *Dictionaries. An International Encyclopedia of Lexicography*. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography. Berlin/Boston: De Gruyter, 517–524.

Kubczak, Jacqueline/Schumacher, Helmut (1998): Verbvalenz – Nominalvalenz. In Bresson, Daniel/Kubczak, Jacqueline (ed.), *Abstrakte Nomina*. *Vorarbeiten zu ihrer Erfassung in einem zweisprachigen syntagmatischen Wörterbuch*. Tübingen: Max Niemeyer, 273–286.

Kühn, Peter (1989): *Typologie der Wörterbücher nach Benutzungsmöglichkeiten.* In: Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (ed.): *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilband. Berlin/New York: De Gruyter (HSK), 111–127.

Lazard, Gilbert (1994): *L'actance*. Paris: Presses Universitaires de France.

Mann, M. (2010): Internetwörterbücher am Ende der "Nullerjahre": Der Stand der Dinge. Eine vergleichende Untersuchung beliebter Angebote hinsichtlich formaler Kriterien unter besonderer Berücksichtigung der Fachlexikographie. In: *Lexicographica* 26, 19–45.

Melchior, Luca (2014): Ansätze zu einer halbkollaborativen Lexikographie. In: *Online publizierte Arbeiten zur Linguistik* 4 (2014), 27–48.

Mel’čuk, Igor (2015): *Semantics. From meaning to text. Volume 3*. Amsterdam/Philadelphia: Benjamin‘s.

Mel’čuk, Igor/Polguère, Alain (2007): *Lexique actif du français*. Louvain la Neuve : De Boeck.

Meyer, Christian/Gurevych, Irina (2012): Wiktionary: A new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography, chapter 13. In: Granger, Sylviane/Paquot, Magali (ed.): *Electronic Lexicography*. Oxford: Oxford University Press, 259–291.

Mirazo Balsa, Mónica (2014): Überlegungen und Vorschläge zur Strukturierung und Darstellung lexikographischer Information in kontrastiven Online-Wörterbüchern. In: Mann, Michael (ed.): *Digitale Lexikographie. Ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen*. Hildesheim/Zürich/New York: Olms, 133–154.

Mirazo Balsa, Mónica (2015): Zur Konzeption und Aufbau eines kontrastiven Substantivvalenzwörterbuches am Beispiel des Forschungsprojektes CSVEA. In: Engelberg, Stefan et al. (ed.): *Argumentstrukturen zwischen Valenz und Konstruktion. Empirie – Theorie – Anwendung*. Tübingen: Narr, 439–449.

Mirazo Balsa, Mónica (2016): El e-diccionario multilingüe de la valencia del sustantivo PORTLEX. Algunas dificultades técnicas y metodológicas en la elaboración de su diseño y estructura. In: Castell, Andreu (2016): *Sintaxis y diccionarios: la complementación en alemán y en español*. Bern: Peter Lang, 89–116.

Model, Benedikt A. (2010): *Syntagmatik im zweisprachigen Wörterbuch*. Berlin: de Gruyter.

Müller-Spitzer, Carolin/Koplenig, Alexander/Töpel, Antje (2012): Online dictionary use: Key findings from an empirical research project. In: Granger, Sylviane / Paquot, Magali (eds.): *Electronic lexicography*. Oxford: Oxford University Press, 425–457.

PORTLEX (2015): *Portlex. Noun phrase mutilingual dictionary. User manual*. [<https://drive.google.com/file/d/0BwAHAUDn37_XOHZkWFE1NUY3QUU/view?usp=sharing>; last access: November 19, 2018].

RAE [Real Academia Española] (2010) = *Nueva gramática de la lengua española*. Manual. Madrid: Espasa Calpe.

Rigau, Gemma (1999): La estructura del sintagma nominal: los modificadores del nombre. In: Bosque, Ignacio/Demonte, Violeta (dirs.): *Gramática descriptiva de la lengua española*. Madrid: Espasa Calpe, 311–362.

Ruppenhofer, Josef et al. (2016): *FrameNet II: Extended Theory and Practice*. Berkeley: International Computer Sciens Institute [<http://https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf>; last access: November 19, 2018].

Schumacher, Helmut (2006a): Deutschsprachige Valenzwörterbücher. In: Ágel, Vilmos et al. (ed.), *Dependenz und Valenz. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin/New York: De Gruyter, vol. 2, 1396–1424.

Schumacher, Helmut (2006a): Kontrastive zweisprachige Valenzwörterbücher. In: Ágel, Vilmos et al. (ed.): *Dependenz und Valenz. Ein internationales Handbuch der zeitgenössischen Forschung*. Berlin/New York: De Gruyter, vol. 2, 1435–1446.

Solla, Miguel Anxo/Gómez, Xavier (2015): Galnet: o WordNet do galego. Aplicacións lexicolóxicas e terminolóxicas. In: *Revista galega de filoloxía* 16, 169–201. [<http://revistas.udc.es/index.php/rgf/article/viewFile/rgf.2015.16.0.1383/628>; last access: October 10, 2017].

Sommerfeldt, Karl-Ernst/Schreiber, Herbert (1977): *Wörterbuch zur Valenz und Distribution der Substantive*. Leipzig: VEB Bibliographisches Institut.

Teubert, Wolfgang (1979): *Valenz des Substantivs. Attributive Ergänzungen und Angaben*. Düsseldorf: Schwann.

Valcárcel Riveiro, Carlos (2017): Las construcciones N_1N_2 como realizaciones actanciales del sustantivo en francés y su tratamiento en el diccionario multilingüe PORTLEX. In: Domínguez Vázquez, Mª José/Kutscher, Silvia (ed.): *Interacción entre gramática, didáctica y lexicografía: estudios contrastivos y multicontrastivos*. Berlin: De Gruyter, 193–207.

Valcárcel Riveiro, Carlos (forthcoming): Lexicografía colaborativa y comunidades digitales en el ámbito románico: el caso de los wikcionarios francés, español y gallego. In: Sánchez Palomino, Mª Dolores et al. (ed.): *Lexicografía iberorrománica*. Madrid: Arco Libros.

9.3 Corpora

CORGA = *Corpus de referencia do galego actual*. Centro Ramón Piñeiro para a Investigación en Humanidades. [<http://corpus.cirp.es/corga>; last access: October 10, 2017].

DeReKo = *Das Deutsche Referenzkorpus*. Institut für Deutsche Sprache. [<http://www1.ids-mannheim.de/kl/projekte/korpora>; last access: October 10, 2017].

CREA = *Corpus de referencia del español actual*. Real Academia Española. [<http://corpus.rae.es/creanet.html>; last access: October 10, 2017].

DEREKO = *Das Deutsche Referenzkorpus*. Institut für Deutsche Sprache. [<http://www1.ids-mannheim.de/kl/projekte/korpora>; last access: October 10, 2017].

FRANTEXT = *Base textuelle FRANTEXT*. ATILF - CNRS & Université de Lorraine. [<http://www.frantext.fr>; last access: October 10, 2017].

PAISÀ = *Piattaforma per l'Apprendimento dell'Italiano Su corpora Annotati*. Università di Bologna/ CNR Pisa/Accademia Europea di Bolzano/Università di Trento. [<http://www.corpusitaliano.it/>; last access: October 10, 2017].

Isabel Durán-Muñoz and Gloria Corpas Pastor

Corpus-based multilingual lexicographic resources for translators: an overview

Abstract: The number of e-resources giving access to electronic corpora is increasing every day, since it is proved that "they provide highly valuable lexical material, patterns, attestations of use, etc." (Leroyer 2015: 257). Following this trend, this paper[1] focuses on translators' needs regarding terminology management and lookup during the different phases of translation tasks, with special reference to some of the main advantages that the combination of corpora and e-resources bring to translation. It also provides an overview of the different types of e-resources that allow access to comparable or parallel corpora in different ways. These e-resources usually employ the Web as Corpus (WaC) (De Schryver 2002), i.e., the web as an unlimited untagged corpus where to consult users' queries about terms or expressions by means of concordance searches. These resources are fairly useful, since they provide users with relevant information about what they are looking for and help them understand some specific information or find tips about how to write or translate something. However, the information offered by these resources is not always reliable, since it is usually the outcome of automatic searches on the Internet, and users are usually not given information about the sources or do not know how to filter that information to select the correct one. Thus, there is a pressing need to conduct terminology research and implement tools specifically addressed at translators. The last part of the paper surveys present-day corpus-based e-resources and provides a tentative classification.

Keywords: terminology look-up, translation, corpus management, term management, e-resource, next-generation intelligent dictionary

[1] This research has been supported in part by the Spanish Ministry of Economy and Competitiveness (grant nos. FFI2012-38881 and FFI2016-75831P), the Andalusian Regional Government (grant no. HUM2754).

Isabel Durán Muñoz: Universidad de Córdoba, Facultad de Filosofía y Letras, Dpto. Filología Inglesa y Alemana, Plaza del Cardenal Salazar, 3, 14071 Córdoba (Spain), tel. +34 957218426, iduran@uco.es
Gloria Corpas Pastor: Universidad de Málaga/University of Wolverhampton, Facultad de Filosofía y Letras, Dpto. Traducción e Interpretación, Avda. Louis Pasteur, 29071 Málaga (Spain), tel. +34 952133409, gcorpas@uma.es

https://doi.org/10.1515/9783110607659-009

1 Introduction

Recent decades have witnessed a constant increase in the rate at which new tools and resources for translation are being developed, motivated by, and fostering at the same time, the gradual technologisation of the profession and translators' habits. Nowadays the translator's workstation usually includes translation memory systems, terminology management systems, term extractors, monolingual/bilingual concordancers, localisation tools and machine translation systems, among many others (cf. LeBlanc 2013). In fact, combinations of some of these tools are often integrated into a single suite, which is increasingly referred to as a Translation Environment Tool (TEnT) (Bowker 2015, Bowker/Corpas Pastor 2015).
Along with the translation process and translators' computerised needs, these systems have also evolved throughout time to adapt themselves and provide more suitable and efficient results in the translation context. As Folaron (2010: 432) states:

> Historically, professional translators have had to find ways to keep up-to-date on the changes in technology produced in response to changing client environments, file formats and language requirements. They have had to become more dexterous and proficient with the tools that perform these functions.

In this vein, there has been a clear transition from desktop to client/server and then to cloud-based services of CAT tools, which eliminate the need for users to own and manage the tool on their computer; from licensed software to subscription business models, whose price is generally based on usage, storage or translation volume on a monthly basis, and toward integrated solutions (cf. TAUS 2013), or TEnT. In fact, as Bowker/Corpas Pastor (2015) state, "TEnTs are the most popular and widely marketed translation tools in use today."

An increase in corpus tools for translation has also been observed in the last decade, mainly as a result of the development of technology and tools regarding compilation, storage and management of corpora[2]. Nowadays, we encounter a myriad of different tools to carry out searches for monolingual/bilingual concordances or collocations (such as *AntCon*[3] or *ParaConc*[4]), alignment of bilingual texts (*YouAlign*[5], *LF Aligner*[6]) or even text analysis online platforms that allow the upload-

[2] For an updated account on uses of corpora in translation (and interpreting), both in professional and academic environments, see Frérot (2016) and the papers included in Corpas Pastor/Seghiri (2016).

[3] [<http://www.antlab.sci.waseda.ac.jp/antconc_index.html>; last access: March 5, 2017].

[4] [<http:// http://paraconc.com/>; last access: March 5, 2017].

[5] [<http://www.youalign.com/>; last access: March 5, 2017].

[6] [<http://lf-aligner.soft112.com/>; last access: March 5, 2017].

ing of both documents and URL addresses and enable users to upload texts or manage linked corpora, such as *Compleat Lexical Tutor*[7] or *TAPoRware*[8]. Even some tools, such as *Sketch Engine*[9], provide an integrated platform enabling users to automatically compile corpora and manage the results in different ways (concordances, collocations, term extraction, etc.), either using part-of-speech tagged or untagged corpora. Moreover, it is common to find some corpora management options (mainly, text alignment and concordance) integrated in both commercial and noncommercial CAT tools, such as *SDL Trados* or *Wordfast Anywhere*. However, there are many other options (term and equivalent extraction, grammar sketches, etc.) which are hardly known or used by translators as some recent surveys have proved (see section 2).

In lexicography, corpora are also frequently employed as an empirical source of information (cf. Fontenelle 2008, Mukherjee 2009). They are seen as a valuable tool to extract relevant data (units, contexts, definitions, pragmatic information, etc.) during the dictionary building stage and, consequently, they are carefully compiled and analysed following specific design protocols according to the project purposes. Consequently, we would expect that information lookup had also evolved according to the current trends. However, once exploited during the dictionary building stage, corpora usually remain hidden and unaccessible to the user. That is, the exploitation phase after the dictionary is completed hardly exists. The truth is that we have witnessed some improvements in the lookup of e-resources and we can now benefit from some of these advantages, such as metasearch engines like *Yourdictionary.com*[10] or *Hispadic*[11] or possible searches for collocations using wildcards and fuzzy matches. There are also tools, such as *InteliWebSearch*[12] or *Terminotix Toolbar*[13], which allow users to customise their searches by enabling them to select several resources at a time and launch a single search in all of them. However, despite these advances, e-resources of today appear to still follow linguistic traditions and practices in the way they present and give access to data and, definitely, they should enter a more advanced dimension in fulfilling more sophisticated needs of their users, for instance if this access to data were not only based on a single lemma (Prinsloo et al. 2011). As a matter of fact, professional translators are still forced to consult a wide range of different resources (glossaries, dictionaries, databases, etc.) when translating in order to find the most suitable options for their translations, or

[7] [<http://www.lextutor.ca>; last access: March 5, 2017].

[8] [<http://taporware.mcmaster.ca>; last access: March 5, 2017].

[9] [<https://www.sketchengine.co.uk/>; last access: March 5, 2017].

[10] [<http://www.yourdictionary.com>; last access: March 5, 2017].

[11] [<http://hispadic.com/>; last access: March 5, 2017].

[12] [<http://www.inteliwebsearch.com>; last access: March 5, 2017].

[13] [<http://www.terminotix.com>; last access: March 5, 2017].

to produce their own smaller but more personalised and high quality terminological resources by using their TEnTs.

In this regard, it goes without saying that dictionaries for translators are needed, as Fuertes-Olivera/Nielsen (2012: 191) state, to provide help to users at various stages of the translation process and combine principles of specialised lexicography and Internet technology. As a way of example, they propose "access to databases through targeted searches", by which lexicographers would present search results in targeted, pre-arranged ways, and would assist the translation of terms, collocations or phrases in direct response to user needs.

It is in this context of e-resources lookup where numerous opportunities to design and develop are open to researchers. Following TAUS (2013: 59): "Accessing up-to-date subject/niche specific terminology is the most common challenge facing people working with languages", and where translators devote most of their time. To this statement, it should be added that offering new ways of consulting the information or finding several contextual options is another challenge that can enhance translation results. In this line, the main goal of this paper is to highlight the need of conducting research addressed at translators as final users underlying the unquestionable advantages that corpora bring to the different phases of any translation task and to present several novel proposals that attempt to fill this existing gap.

The rest of this chapter is organised as follows. The first part focuses on translators' needs regarding terminology management and lookup, beginning with an overview of the terminology impact on the different phases of translation tasks and followed by a more detailed look at some of the main advantages that corpora bring to translation. The second part of the paper offers an overview of a number of e-resources that provide the possibility to access comparable or parallel corpora in different ways. The closing section is devoted to some final thoughts concerning the combination of e-resources and corpora and the positive aspects that this can bring to translation.

2 Specialised lexicography in the translation process: the role of corpora

Translators' need for special knowledge plays an important role in all stages of the translation process: to ascertain the meaning of a term in the source language, to find equivalents for the target text, to select the most suitable option among several alternative specialised units and, of course, to acquire knowledge about a specialised domain by means of their knowledge structures (cf. Cabré Castellví 2010: 358). As such, the three different stages of any translation task present different needs:

- In the pre-translation stage, when translators first approach and try to understand the source text by learning the meaning of specialised units and/or their grammatical and pragmatic conditions of use;
- during the translation stage itself, when the source text is written in the target text and mainly equivalence problems are encountered, i.e. how to find an equivalent unit or how to select the most appropriate equivalent among several options; and
- in the post-translation stage, when translators carry out the revision of the draft and produce their final version by controlling textual coherence, inconsistencies and quality of the final output.

Specialised knowledge as well as specialised units in a given field are, then, key to the whole translation process. Recent advances have acknowledged this fact by producing tools that enhance information lookup and multiple searches (see section 3.4). However, this is not enough and most of the existing applications seem to be of limited use, as translators are still forced to consult a plethora of resources during the translation task. According to Lew (2011: 241–242), the retrieved entries are often very similar and this “results in highly unhelpful, many times redundant, tortuous assemblages of disconnected lexicographic data”. In some cases, these resources do not even resolve translators’ doubts and they need to apply other strategies by carrying out labour-intensive and fruitless searches to find answers and, then, produce their own resources, mainly within their TEnT tools. These unproductive searches are frequently caused by the lack of updated resources or their inadequacy for the specific translators’ needs (Cabré Castellví 2010: 360).

In addition, the exponential growth of information and knowledge of specialised domains prompts to the continuous appearance of newly coined specialised units, particularly in some fields where term dynamics is very high (cf. Temmerman/van Campenhoudt 2014). Given the novelty of those terms, it is impossible to collect them all and keep specialised resources updated. Indeed, these new units are one of the most common difficulties found by translators and lexicographers, which greatly hinders the quality of their outputs. Corpora, and its management by means of concordance, collocation, term extraction or alignment tools, are a very suitable solution for those tedious and unproductive searches, as well as to provide access to new specialised units that are not yet included in resources (cf. Zanettin 2012, Zanettin et al. 2015, Kruger et al. 2011, Beeby et al. 2009, Olohan 2004, Kruger et al. 2011, etc.).

There is a wide variety of corpora[14] that prove very useful for different purposes in translation and terminology tasks. Apart from reducing the fruitless and exhaust-

[14] For further information regarding types of corpora, please consult EAGLES (1994, 1996), Corpas Pastor (2001: 157–159, 2004: 225–228, 2008: 117–118) and Bowker/Pearson (2002: 12–13).

ing searches which translators are obliged to when they do not easily find information in lexicographical resources, corpora are of great use when dealing with pragmatic information. Dictionaries and other lexicographical resources in general usually offer context-free descriptions of word use, which is not enough for translators. Translators and, in general, linguistic mediators require equivalents and linguistic information (included in dictionaries and alike), but they also need reassurance regarding real use in context. That is, they need to be completely sure that the unit or expression they are employing in a specific target text is the best option to translate the source term or expression, not only regarding meaning but also register, style, geographical variant, etc. For those reasons, corpora are helpful when checking geographical variations and different spellings, finding collocations and frequent word associations, testing idiomatic phrases, and detecting typical, recurrent and repeatedly observable contexts in specific registers, among other possibilities. This information is rarely included in terminological resources, and this is mainly the reason why corpora are of such a great value for translators.

In the following lines we summarise the advantages that corpora bring to the different phases of translation:

- Comparable corpora of original texts (be monolingual, bilingual or multilingual) are suitable resources for the acquisition of specialised knowledge, especially at the pre-translation phase when translators approach the original texts for the first time and need to grasp the original meaning. At this point, corpora can provide valuable information by helping translators build and understand the conceptual map of the specialised domain. Translators can use term extractors in both working languages and analyse the most frequent terms that occur in both languages, or they can align parallel texts to easily observe the original and the translated segment (usually at the paragraph level) and use the aligned texts as a source to create their own translation memories.
- During the translation phase, concordancers and collocates can provide relevant information about the real use of the searched terms and expressions, ranging from giving information about geographical variations and spelling to different registers. Also, they can provide information about the use and meaning of developed forms of acronyms and abbreviations, use of synonyms in different contexts, such as patient vs. subject or irregular heart rhythm vs. arrhythmia, as well as grammatical preferences, such as worth to do vs. worth doing or even serendipitous findings that can be very valuable sometimes.
- At the revision phase, corpora can also serve as tools to check terminology spellings, abbreviations, acronyms, full forms, etc. and, therefore, assist translators and correctors to revise the final text before completing the process and provide high quality in their translations and/or revisions.

In light of the above, it can be established that corpora are of great importance for translators (both as empirical sources of linguistic and pragmatic information) and, thus, they are considered as essential tools in their profession. However, few lexicographical resources include corpora among their options and few professional translators employ (or even know about) corpora in their translation tasks, as several studies carried out on translation technologies have shown. By way of example, we refer to a survey conducted in the framework of the FP7 project TTC (Terminology Extraction, Translation Tools and Comparable Corpora (ICT-2009-248005) about terminology and corpus practices (Gornostay 2010, Blancafort et al. 2011). This survey was answered by 139 language professionals from thirty-one countries and interested in the use of corpora, corpus tools, and NLP applications, among others. The results shed some light on professional translators' preferences and uses on corpora: half of the respondents collected corpora of their working fields but only 7% of them employed concordancers and other NLP tools (mainly POS taggers) for corpus processing. They even indicated their preferences regarding manual search for relevant terms and equivalents. As corpus compilation and processing is concerned, the conclusions of this survey were that corpus compilation was time-consuming and corpus tools were largely unknown. Those results are in line with Durán-Muñoz (2012), whose survey indicated that professional translators tend to compile *ad hoc* corpora when translating (65.21% of the respondents)[15], in order to check terms in context and to extract terms to populate their own termbanks, NLP tools are not mentioned at all and only 14.29% seem to use some kind of corpus management and processing tools (e.g. WordSmith Tools).

A more recent survey (Zaretskaya et al. 2018) confirms previous findings. The purpose of this survey was to ascertain the degree of technologisation in the professional translators' workflow, including corpora and their processing. According to the respondents[16], tools for compiling or managing corpora were the least commonly used on a regular basis (only 4% of total participants). In the same vein, concordance tools were completely unknown to the majority of respondents, and term extraction tools were employed by only 25% of participants.

These surveys prove that corpora are still largely unknown among professional translators, despite the clear advantages that they bring when dealing with terminology and specialised knowledge. As stated above, the purpose of this paper is to shed some light of these unquestionable advantages and outline specific corpus-based lexicographical e-resources that assist translators during their work, which are presented in the next sections.

[15] The total number of respondents participating in this survey was 402, working in different countries and languages.

[16] According to the authors, there were 736 completed responses from 88 different countries, and the majority of respondents were experienced translators and almost a half of them had more than 11 years in the industry.

3 Corpus-based multilingual e-resources for knowledge information tasks

As previously mentioned, many translators are still not aware of the advantages that corpora bring to their work or are reluctant to employ them. Nevertheless, there are more and more e-resources in the market that provide direct access to corpora by means of corpus management tools.

In the next subsections, we focus on these e-resources, paying special attention to multilingual e-resources, and provide an overview of the different types and main features by classifying them in four different groups.

3.1 Free, web-searchable online corpora

Free, web-searchable corpora are growing in number, particularly in the last years, due to the improvement of storage, speed of access and response, etc. of computer systems. Some of these corpora have been compiled for a long time now but it is just recently that they have become accessible and manageable. They are usually the result of concise research carried out within public and private institutions and, as such, they offer a very high level of reliability.

They are very large corpora, over millions of words, and they are usually monolingual: Corpus of Contemporary American English (COCA), British National Corpus (BNC), Reference Corpus of Contemporary Spanish (CREA), Corpora Linguistici per l'Italiano Parlato e Scritto (CLIPS), etc. Some of them may restrict the number of searches per day, like COCA or BNC, and most of them also provide the possibility to customise the searches by selecting different textual genres, periods of time, among other criteria.

There are also multilingual corpora of this kind, however they are usually built in the framework of research projects and, thus, accessible with limited usage or not accessible at all. Examples of these types of corpora are, for instance, the Oslo Multilingual Corpus[17], the European Corpus Initiative (ECI)[18], or Multi-CAST: Multilingual Corpus of Annotated Spoken Texts[19].

These are considered very reliable resources, as they have been compiled in the framework of research projects or public institutions. However, when dealing with specialised translation, it is important to bear in mind that they are usually general corpora and, thus, they are not domain-specific. In other words, the information

[17] [<https://www.hf.uio.no/ilos/english/services/omc/>; last access: March 5, 2017].

[18] [<http://www.elsnet.org/eci.html/>; last access: March 5, 2017].

[19] [<https://www.uni-bamberg.de/en/aspra/research/projects/multi-cast-multilingual-corpus-of-annotated-spoken-texts//>; last access: March 5, 2017].

about specialised units that can be consulted is limited. Another drawback is that there is a lack of homogeneity when managing these corpora, since they are stored and managed by different systems, which offer diverse options of access and consultation according to the system features and the corpora pre-processing (annotated or not annotated corpora).

To sum up, this type of corpora provides a wide range of linguistic and pragmatic information which is very useful for translators, especially when writing in their non-native language. They can check collocations, concordances, as well as general usage and style. However, there are some disadvantages that still hamper the spread of these corpora among translators, particularly due to the fact of being too general to retrieve adequate and accurate results when translating specialised texts. In addition, these corpora are usually monolingual and, less frequently, bilingual corpora. Hence, their use for domain-specific translation is limited.

3.2 Corpus-based web crawlers

Corpus-based web crawlers are defined as the tools that employ the Internet (the Web) as a direct source of information to launch linguistic queries or compile corpora automatically, managing the *Web as Corpus* (WaC) (De Schryver 2002)[20], i.e., the web as an unlimited untagged corpus where to consult users' queries.

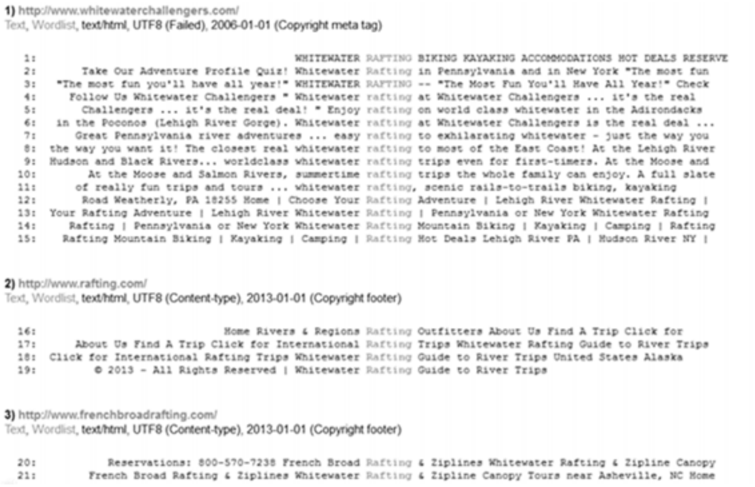
Due to space constraints, this paper will only present tools that allow linguistic queries directly on the Web[21], which can be classified in two main groups:

- Search engines, such as Google, and their advanced search. It allows users to customise and launch multilingual searches and delimit them by introducing or selecting specific information in the available fields (country, period of time, document format, etc.), together with the working languages. This type of searches results in a list of websites which meet the requirements previously specified by the users and, thus, include information relevant to them.
- Web concordancers. These tools are concordancers, that is, tools that look for keywords in context (KWIC) but, instead of searching on a specific corpus, they do it on the whole Web. The queries introduced by the users are launched to the Web and, as a result, the tool displays a list of concordance lines, which include the term or expression searched by the user and the words that accompany

[20] De Schryver (2002: 272) distinguishes between the Web as Corpus, in which the focus in "on the potential of the Web 'as' a corpus in itself" or the Web for Corpus, i.e., the Web "as a provider of data 'for' the creation of corpora."

[21] Other corpus-based web crawlers that are included in this group are automatic corpus compilation tools, such as BootCat [http://bootcat.sslmit.unibo.it/>; last access: March 5, 2018] and WebBootCat - Sketch Engine [<http://www.sketchengine.co.uk/>; last access: March 5, 2018].

those searched term/expressions (right and left co-text) (cf. Figure 5). As examples of web concordancers, we can name WebCorp[22] and Wordincontext[23].



```
1) http://www.whitewaterchallengers.com/
Text, Wordlist, text/html, UTF8 (Failed), 2006-01-01 (Copyright meta tag)

 1:                                       WHITEWATER RAFTING BIKING KAYAKING ACCOMMODATIONS HOT DEALS RESERVE
 2:      Take Our Adventure Profile Quiz! Whitewater Rafting in Pennsylvania and in New York "The most fun
 3:  "The most fun you'll have all year!" WHITEWATER RAFTING -- "The Most Fun You'll Have All Year!" Check
 4:    Follow Us Whitewater Challengers " Whitewater rafting at Whitewater Challengers ... it's the real
 5:      Challengers ... it's the real deal! " Enjoy rafting on world class whitewater in the Adirondacks
 6:  in the Poconos (Lehigh River Gorge). Whitewater rafting at Whitewater Challengers is the real deal ...
 7:     Great Pennsylvania river adventures ... easy rafting to exhilarating whitewater - just the way you
 8: the way you want it! The closest real whitewater rafting to most of the East Coast! At the Lehigh River
 9: Hudson and Black Rivers... worldclass whitewater rafting trips even for first-timers. At the Moose and
10:       At the Moose and Salmon Rivers, summertime rafting trips the whole family can enjoy. A full slate
11:     of really fun trips and tours ... whitewater rafting, scenic rails-to-trails biking, kayaking
12:      Road Weatherly, PA 18255 Home | Choose Your Rafting Adventure | Lehigh River Whitewater Rafting |
13: Your Rafting Adventure | Lehigh River Whitewater Rafting | Pennsylvania or New York Whitewater Rafting
14:    Rafting | Pennsylvania or New York Whitewater Rafting Mountain Biking | Kayaking | Camping | Rafting
15:   Rafting Mountain Biking | Kayaking | Camping | Rafting Hot Deals Lehigh River PA | Hudson River NY |

2) http://www.rafting.com/
Text, Wordlist, text/html, UTF8 (Content-type), 2013-01-01 (Copyright footer)

16:                            Home Rivers & Regions Rafting Outfitters About Us Find A Trip Click for
17:     About Us Find A Trip Click for International Rafting Trips Whitewater Rafting Guide to River Trips
18: Click for International Rafting Trips Whitewater Rafting Guide to River Trips United States Alaska
19:        © 2013 - All Rights Reserved | Whitewater Rafting Guide to River Trips

3) http://www.frenchbroadrafting.com/
Text, Wordlist, text/html, UTF8 (Content-type), 2013-01-01 (Copyright footer)

20:          Reservations: 800-570-7238 French Broad Rafting & Ziplines Whitewater Rafting & Zipline Canopy
21:       French Broad Rafting & Ziplines Whitewater Rafting & Zipline Canopy Tours near Asheville, NC Home
```

Fig. 1: Concordance lines of the search "rafting" in *WebCorp*

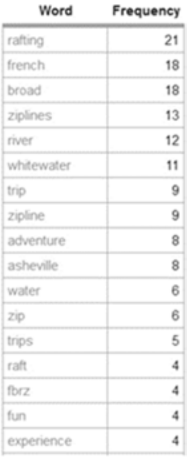
Users are given the opportunity to carry out multilingual searches in different languages. *WebCorp*, for example, works for English, German and Chinese; and *WordinContext* offers a long list of working languages (18 languages in total), including Spanish, English, Italian, Portuguese, English, German and French, along with other less common languages, such as Esperanto, Suomi, Catalan or Polish.

The results in these e-resources are usually organised by websites and the keyword (searched term or expression) appears highlighted for the sake of clarity and efficiency, as it can be observed in Figure 1. Some of these e-resources also offer other options, such as a word list with the most frequency words of the selected websites (Figure 2).

These web-crawlers are very easy to use and intuitive. However, when using them, it is important to bear in mind one of their main disadvantages (in general the Web as corpus tools) which is the so-called *noise* (or irrelevant information) they generate during their searches. This requires a thorough post-revision of results by users.

[22] [<http://www.webcorp.org.uk/>; last access: March 5, 2018].

[23] [<http://wordincontext.com/>; last access: March 5, 2018].



Word	Frequency
rafting	21
french	18
broad	18
ziplines	13
river	12
whitewater	11
trip	9
zipline	9
adventure	8
asheville	8
water	6
zip	6
trips	5
raft	4
fbrz	4
fun	4
experience	4

Fig. 2: Word list generated by *WebCorp* in the search "rafting"

3.3 Web-based translation memories

Translation memory (TM) systems are one of the most widely used technologies by professional translators and translation companies (Garcia 2007, Christensen/ Schjoldager 2010, O'Hagan 2013, Bowker/Corpas Pastor 2015). There is a wide range of systems, both commercial and freeware, desktop or web-based, but they all share the core system functionality: to allow translators to access a database containing aligned texts (translated texts and their source texts) while translating. As such, the system suggests, at the moment of translating, possible translations which have been previously used as equivalents to similar source segments[24] and which are stored in the database. These systems can be populated either by adding previously aligned translated texts and their corresponding source texts, i.e., previous translation memories, or by starting to translate from scratch. In the latter case, the TM will be empty at the beginning and will be populated by the user.

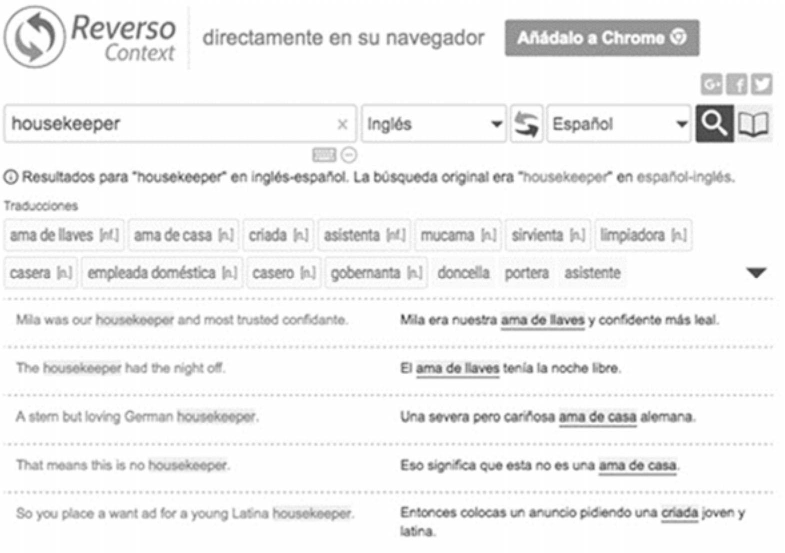
For its part, web-based TM systems perform in a similar way as the systems aforementioned: they also use aligned translated and source texts as their source in a long list of working languages, but they present several differences:

– They are not integrated in computer-assisted translation (CAT) tools.
– They are accessible via the Internet and, thus from anywhere and at anytime.
– They use the web as a parallel corpus to find source and target texts in bilingual or multilingual websites.

[24] In these systems, segments are usually complete sentences or paragraphs, depending on the TM system.

- Users have to launch specific queries by using their search option.
- They offer a great number of language combinations.

These web-based systems offer as result a list of sentences (or fragments) which include the looked-up unit or expression and the equivalent sentence (or fragment) in the selected languages, which have been extracted from bilingual or multilingual websites (see Figure 3). Some of the most popular web-based translation memories are *Linguee*[25], *MyMemory*[26] and *Reverso Context*[27].



ⓘ Resultados para "housekeeper" en inglés-español. La búsqueda original era "housekeeper" en español-inglés.

Traducciones

ama de llaves [nf.] | ama de casa [n.] | criada [n.] | asistenta [nf.] | mucama [n.] | sirvienta [n.] | limpiadora [n.] | casera [n.] | empleada doméstica [n.] | casero [n.] | gobernanta [n.] | doncella | portera | asistente

Mila was our housekeeper and most trusted confidante.	Mila era nuestra ama de llaves y confidente más leal.
The housekeeper had the night off.	El ama de llaves tenía la noche libre.
A stern but loving German housekeeper.	Una severa pero cariñosa ama de casa alemana.
That means this is no housekeeper.	Eso significa que esta no es una ama de casa.
So you place a want ad for a young Latina housekeeper.	Entonces colocas un anuncio pidiendo una criada joven y latina.

Fig. 3: Results of "housekeeper" in *Reverso Context*

They are frequently employed by professional and trainee translators as they are very useful tools. They provide equivalents in a very easy and quick way in a wide range of language combinations, both original and target language. However, they also need to be treated with caution. These systems provide the results extracted by bilingual and multilingual websites and they usually employ reliable sources, such as public and private institutions, international organisations, etc. but the translations are not classified according to their level of specialisation, geographical variation, accuracy of the equivalent, etc. and some of them do not provide information

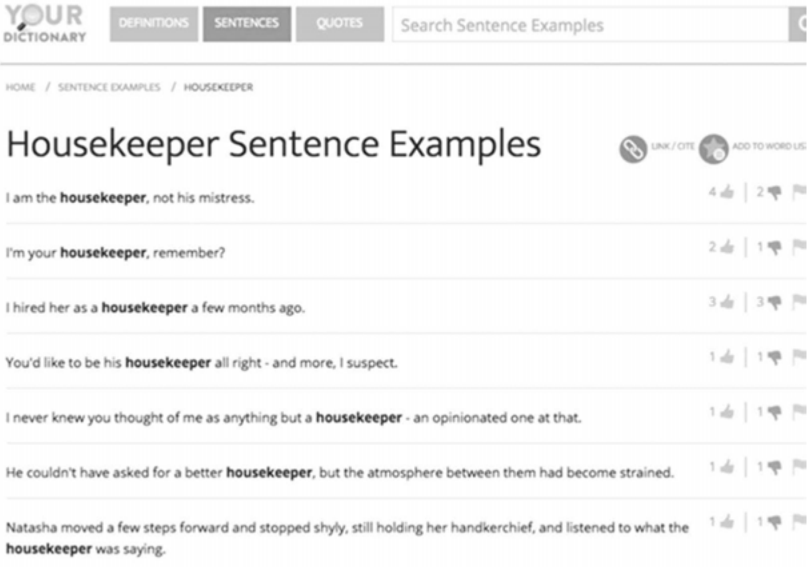
[25] [<http://www.linguee.com>; last access: March 5, 2017].
[26] [<https://mymemory.translated.net/>; last access: March 5, 2017].
[27] [<http://context.reverso.net/>; last access: March 5, 2017].

about the sources. Besides, they often offer a great variety of equivalents of the same source term or expression which lead to confusion and uncertainty.

Consequently, the reliability and specialisation of these e-resources are frequently questionable and, hence, they must be regarded as useful e-resources that provide a starting point to translators.

3.4 Combination of corpus and e-dictionaries

The number of e-dictionaries that offer access to corpora is increasing. These e-resources provide users with relevant information about what they are looking for and help them understand some specific information or find tips about how to write or translate something. This type of e-resources show contextual examples by means of concordance lines. Some relevant examples are *Glosbe*[28], *Bab.la Dictionary*[29], *Dictionary.com*[30] and *Your Dictionary*[31].



Housekeeper Sentence Examples

I am the **housekeeper**, not his mistress. 4 | 2

I'm your **housekeeper**, remember? 2 | 1

I hired her as a **housekeeper** a few months ago. 3 | 3

You'd like to be his **housekeeper** all right - and more, I suspect. 1 | 1

I never knew you thought of me as anything but a **housekeeper** - an opinionated one at that. 1 | 1

He couldn't have asked for a better **housekeeper**, but the atmosphere between them had become strained. 1 | 1

Natasha moved a few steps forward and stopped shyly, still holding her handkerchief, and listened to what the **housekeeper** was saying. 1 | 1

Fig. 4: Results of “housekeeper” in *Your Dictionary*

[28] [<https://es.glosbe.com/>; last access: March 5, 2018].
[29] [<http://en.bab.la/dictionary>; last access: March 5, 2018].
[30] [<http://www.dictionary.com/browse/context/>; last access: March 5, 2018].
[31] [<http://www.yourdictionary.com/>; last access: March 5, 2018].

These e-dictionaries also employ the *Web as Corpus* and they usually do not filter the information that they are displaying to their users as final results. This information is not always reliable as the outcome is usually the product of automatic searches on the Internet, and users are frequently not given information about the sources or do not know how to filter that information to select the correct one.

Due to the drawbacks of these tools regarding unreliable and/or unknown information sources and following a steady upward trend of integrating corpora to e-dictionaries (cf. Héja 2010, Hanks 2012, Fuertes-Olivera/Tarp 2014), there are some current research projects under development that attempt to fill this existing gap[32]. In line with this, *Inteliterm*[33] is an intelligent modular system that combines an efficient lookup dictionary that includes several types of searches (contextual, cross-reference, external searches), as well as corpus management tools. *Inteliterm* covers the domain of health and beauty tourism in four languages (Spanish, English, German and Italian). The main aim of this tool is to meet translators' needs and reduce the fruitless time devoted to knowledge information tasks. This multilingual e-dictionary offers not only access to concordances, as the other e-resources aforementioned, but also the other main uses of corpora for translation, namely related terms (derivatives and other formal variations), cognates, n-grams and parallel texts (cf. Durán-Muñoz et al. 2015, see Corpas Pastor/Durán-Muñoz, this volume).

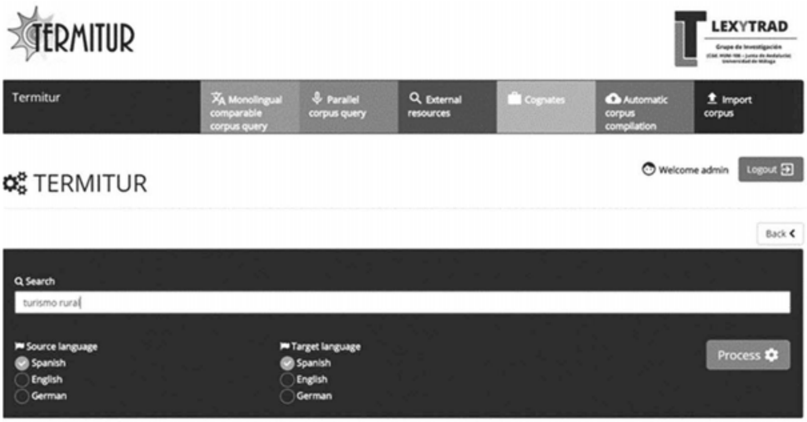
Another interesting tool of this kind is *Termitur*[34], which is currently under development. It follows in the footsteps of the INTELITERM project. As in the preceding project, the main goals are to contribute to the development of systems which facilitate application and technological transfer in order to formulate, implement and evaluate new management options and knowledge acquisition. However, *Termitur* is not only a database, but an integrated and flexible tool that enables users to (semi)automatically compile a corpus or import their own corpus in order to make use of it through the different modules, providing a more complex experience and facilitating the work to linguists, terminologists, translators, etc.

[32] On recent developments in the field of electronic lexicography, the integration of corpora and e-dictionaries, as well as some innovative dictionary projects, see the papers in Granger and Paquot (2012). Dutsova (2015) briefly describes a general-purpose web-based digital bilingual resource that combines two sets of natural language data (bilingual dictionary and aligned text corpora).

[33] *Inteliterm* is a web-based application that has been designed and developed within the framework of the Spanish R&D project INTELITERM: Sistema inteligente de gestión terminológica para traductores [INTELITERM: Intelligent Terminology Management System for Translators] (ref. no. FFI2012-38881, 2012-2016). See Corpas Pastor/Durán-Muñoz (2018).

[34] *Termitur* is a web-based application that has been designed and developed within the framework of the Regional R&D project TERMITUR: Diccionario inteligente TERMInológico para el sector TURístico (alemán-inglés-español) [TERMITUR: Intelligent Terminology Dictionary for the tourism sector (German-English-Spanish] (Ref. HUM2754, 2014-2017. Junta de Andalucía).

Termitur is a next-generation multilingual lexicographical tool. The main dictionary module is intertwined with other modules: terminology management, resources lookup, (semi)automatic corpus compilation and corpus management, both comparable and parallel. The resulting system enables users to upload dictionaries, glossaries and corpora or, else, to compile corpora automatically according to their needs. Further functionalities include identification of cognates, collocation extraction, multiword query search strings, monolingual and bilingual concordances, etc.

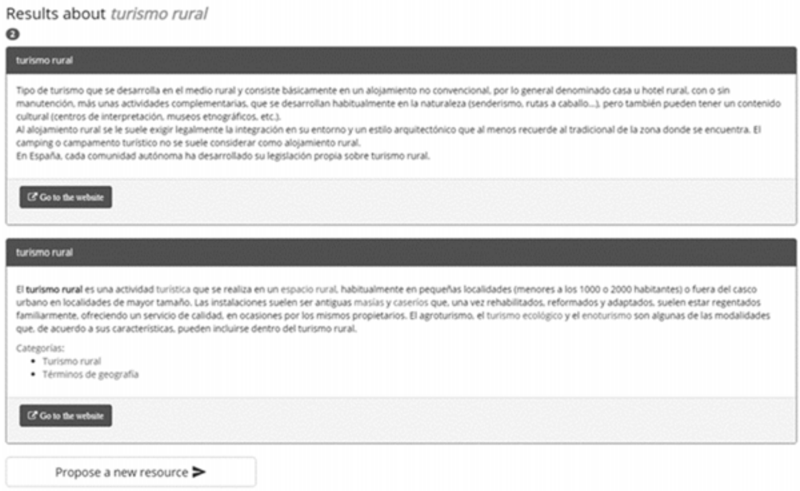


Search
turismo rural

Source language
● Spanish
○ English
○ German

Target language
● Spanish
○ English
○ German

Fig. 5: Search query for "turismo rural" in *Termitur*


Results about *turismo rural*

2

turismo rural

Tipo de turismo que se desarrolla en el medio rural y consiste básicamente en un alojamiento no convencional, por lo general denominado casa u hotel rural, con o sin manutención, más unas actividades complementarias, que se desarrollan habitualmente en la naturaleza (senderismo, rutas a caballo...), pero también pueden tener un contenido cultural (centros de interpretación, museos etnográficos, etc.).
Al alojamiento rural se le suele exigir legalmente la integración en su entorno y un estilo arquitectónico que al menos recuerde al tradicional de la zona donde se encuentra. El camping o campamento turístico no se suele considerar como alojamiento rural.
En España, cada comunidad autónoma ha desarrollado su legislación propia sobre turismo rural.

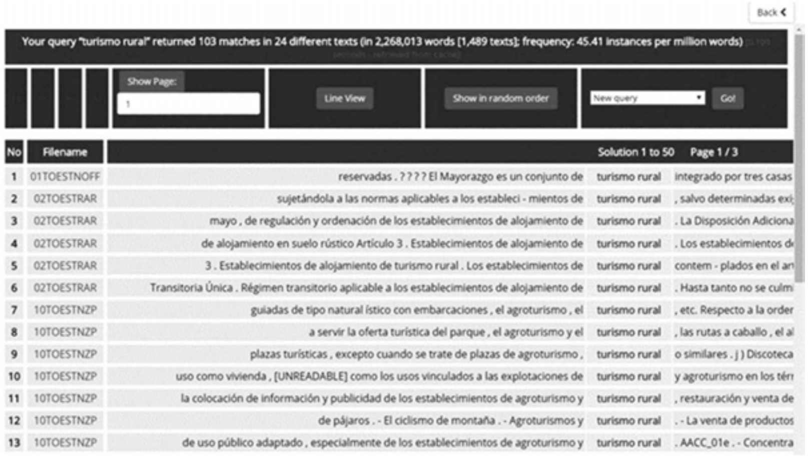
turismo rural

El **turismo rural** es una actividad turística que se realiza en un espacio rural, habitualmente en pequeñas localidades (menores a los 1000 o 2000 habitantes) o fuera del casco urbano en localidades de mayor tamaño. Las instalaciones suelen ser antiguas masías y caseríos que, una vez rehabilitados, reformados y adaptados, suelen estar regentados familiarmente, ofreciendo un servicio de calidad, en ocasiones por los mismos propietarios. El agroturismo, el turismo ecológico y el enoturismo son algunas de las modalidades que, de acuerdo a sus características, pueden incluirse dentro del turismo rural.

Categorías:
- Turismo rural
- Términos de geografía

Fig. 6: *Termitur* results page (External resources)

The final result is a hybrid corpus-based e-resource that allows translators and interpreters to acquire specialised knowledge on the domain of rural and natural tourism. It covers German, English and Spanish, as well as in the resulting language pairs. *Termitur* allows users to perform a simultaneous search in all the modules, i.e. external resources, cognates, a monolingual comparable corpus management system, and a parallel corpus management system, by typing the desired term in the search field. The search is performed automatically and the results obtained are presented in a fast and compact way. Figure 5 illustrates a sample query string (ES *turismo rural*, 'rural tourism') in *Termitur*. Figures 6 and 7 show results for this search query in the modules of External resources and Monolingual comparable corpus query.



Your query "turismo rural" returned 103 matches in 24 different texts (in 2,268,013 words [1,489 texts]; frequency: 45.41 instances per million words)

No	Filename		Solution 1 to 50	Page 1 / 3
1	01TOESTNOFF	reservadas . ? ? ? ? El Mayorazgo es un conjunto de	turismo rural	integrado por tres casas
2	02TOESTRAR	sujetándola a las normas aplicables a los estableci - mientos de	turismo rural	, salvo determinadas exi
3	02TOESTRAR	mayo , de regulación y ordenación de los establecimientos de alojamiento de	turismo rural	. La Disposición Adiciona
4	02TOESTRAR	de alojamiento en suelo rústico Artículo 3 . Establecimientos de alojamiento de	turismo rural	. Los establecimientos de
5	02TOESTRAR	3 . Establecimientos de alojamiento de turismo rural . Los establecimientos de	turismo rural	contem - plados en el ar
6	02TOESTRAR	Transitoria Única . Régimen transitorio aplicable a los establecimientos de alojamiento de	turismo rural	. Hasta tanto no se culm
7	10TOESTNZP	guiadas de tipo natural ístico con embarcaciones , el agroturismo , el	turismo rural	, etc. Respecto a la order
8	10TOESTNZP	a servir la oferta turística del parque , el agroturismo y el	turismo rural	, las rutas a caballo , el al
9	10TOESTNZP	plazas turísticas , excepto cuando se trate de plazas de agroturismo ,	turismo rural	o similares . j) Discoteca
10	10TOESTNZP	uso como vivienda , [UNREADABLE] como los usos vinculados a las explotaciones de	turismo rural	y agroturismo en los térr
11	10TOESTNZP	la colocación de información y publicidad de los establecimientos de agroturismo y	turismo rural	, restauración y venta de
12	10TOESTNZP	de pájaros . - El ciclismo de montaña . - Agroturismos y	turismo rural	. - La venta de productos
13	10TOESTNZP	de uso público adaptado , especialmente de los establecimientos de agroturismo y	turismo rural	. AACC_01e . - Concentra

Fig. 7: *Termitur* results page (Monolingual corpus)

4 Conclusions

Corpora are still largely unknown among professional translators (cf. Gallego-Hernández 2015, Picton et al. 2015, Frankenberg-Garcia 2015), despite the advantages that they bring when dealing with specialised knowledge. And therefore, more training, practical knowledge and applications are needed if corpora are to be extensively used by professional translators.

Judging from the advantages brought about by corpora in translation and terminology work, it would be certainly most beneficial that researchers and developers worked together along those lines in order to propose new integrated tools that overcome the limitations identified to date and include corpora in their applications. In this paper, we present an overview of the different possibilities and the advantages and possible improvements that can be useful for future research and serve as the basis for developing new and complete e-resources.

As a final remark, it is important to emphasise the current trend of developing e-resources that enable corpus access to users, mainly based on the *Web as Corpus* concept, and the scarcity of e-dictionaries that offer this option. Few e-dictionaries do, and those usually are based on the *Web as Corpus* concept. However, specialised translation and interpreting need accurate terminology resources that cover specific fields and domains, not just general reference. In this respect, next-generation multilingual e-resources are urgently needed. It is of paramount importance, then, that lexicographers recognise this existing gap and work accordingly in order to develop high quality e-resources that also allow end-users to consult and manage corpora in an effective and easy way.

5 Bibliography

Beeby, Allison/Rodríguez Inés, Patricia/Sánchez-Gijón, Pilar (2009): Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate. Amsterdam: Benjamin's.

Blancafort, Helena/Heid, Ulrich/Gornostay, Tatiana/Méchoulam, Claude/Daille, Béatrice/Sharoff, Serge (2011): User-centred Views on Terminology Extraction Tools: Usage Scenarios and Integration into MT and CAT Tools. In: Proceedings TRALOGY Conference "Translation, Careers and Technologies: Convergence Points for the Future. Paris: Institut de l'Information Scientifique et Technique [<https://hal.archives-ouvertes.fr/hal-00818657/document>; last access: November 19, 2018].

Bowker, Lynne (2015): Broad approaches to terminology management in a translation context. In: Hendrik J. Kockaert/Frieda Steurs (eds.): Handbook of Terminology Online. Amsterdam/Philadelphia: Benjamin's.

Bowker, Lynne/Corpas Pastor, Gloria (2015): Translation Technology. In: Mitkov, Ruslan (ed.): The Oxford Handbook of Computational Linguistics. 2nd edition. Oxford: Oxford University Press.

Bowker, Lynne/Pearson, Jennifer (2002): Working with Specialized Language: A practical guide to using corpora. London: Routledge.

Cabré Castellví, M. Teresa (2010): Terminology and translation. In: Gambier, Yves/Van Doorslaer, Luc (eds.): Handbook of translation studies, vol.1. Amsterdam: Benjamin's, 356–365.

Corpas Pastor, Gloria (2001): Compilación de un corpus ad hoc para la enseñanza de la traducción inversa. In: TRANS 5, 155–179.

Corpas Pastor, Gloria (2004): Localización de recursos y compilación de corpus vía Internet: aplicaciones para la didáctica de la traducción médica especializada. In: Gonzalo García, Consuelo/García Yebra, Valentín. (eds.): Manual de documentación y terminología para la traducción especializada. Madrid: Arco Libros S.L., 223–258.

Corpas Pastor, Gloria (2008): Investigar con corpus en traducción: los retos de un nuevo paradigma. Frankfurt am Main: Peter Lang.

Corpas Pastor, Gloria/Durán-Muñoz, Isabel (2018): Inteliterm: in search of efficient terminology lookup tools for translators. In: Domínguez, Mª José/Mirazo, Mónica/Valcárcel, Carlos (eds.): Studies on Multilingual Lexicography. Berlin/Boston: De Gruyter, 115-134.

Corpas Pastor, Gloria/Seghiri, Míriam (eds.) (2016): Corpus-based Approaches to Translation and Interpreting. From Theory to Applications. Frankfurt am Main: Peter Lang.

Christensen, Tina Paulsen/Anne Schjoldager (2010): Translation-Memory (TM) Research: What Do We Know and How Do We Know It? In: Hermes—The Journal of Language and Communication 44, 1–13.

De Schryver, Gilles Maurice (2002): State-of-the-Art Software to Support Intelligent Lexicography. In: Zhu, Ruiping (ed.): Proceedings of the International Seminar on Kangxi Dictionary & Lexicology. Beijing: Beijing Normal University, 565–580.

Durán-Muñoz, Isabel (2012): La ontoterminografía aplicada a la traducción: Propuesta metodológica para la elaboración de recursos terminológicos dirigidos a traductores. Studien zur romanischen Sprachwissenschaft und interkulturellen Kommunikation 80. Frankfurt am Main/New York: Peter Lang.

Durán Muñoz, Isabel/Fernández Sola, Alejandro/Corpas Pastor, Gloria (2015): INTELITERM: Una herramienta innovadora para la gestión de corpus y textos para la traducción especializada. In: Corpas Pastor, Gloria/Seghiri Domínguez, Míriam/Gutiérrez Florido, Rut/Urbano Mendaña, Míriam (eds.): Nuevos horizontes en los Estudios de Traducción e Interpretación. Geneve: Tradulex, 248–458.

Dutsova, Ralitsa (2015): Web-based digital lexicographic bilingual resources. In: Cognitive Studies/Études cognitives 15, 369–377.

EAGLES (1994) = Corpus Typology: A framework for classification. EAGLES Technical document 080294, 1–18.

EAGLES (1996) = Preliminary Recommendations on Corpus Typology. EAGLES Technical document EAG-TCWG-CTYP/P.

Folaron, Debbie (2010): Translation Tools. In: Gambier, Yves/van Doorslaer, Luc (eds.): Handbook of Translation Studies, volume 1. Amsterdam/Philadelphia: Benjamin's, 429–436.

Fontenelle, Thierry (ed.) (2008): Practical Lexicography: A Reader. Oxford: Oxford University Press.

Frankenberg-Garcia, Ana (2015): Training translators to use corpora hands-on: challenges and reactions by a group of 13 students at a UK university. In: Corpora 10 (2), 351–380.

Frérot, Cécile (2016): Corpora and Corpus Technology for Translation Purposes in Professional and Academic Environments. Major Achievements and New Perspectives. In: Cadernos de Tradução 36, 36–61.

Fuertes-Olivera, Pedro A./Nielsen, Sandro (2012): Online Dictionaries for Assisting Translators of LSP Texts: The Accounting Dictionaries. International Journal of Lexicography 25 (2), 191–215.

Fuertes-Olivera, Pedro A./Tarp, Sven (2014): Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography. (Lexicographica. Series Maior 146). Berlin/Boston: De Gruyter.

Gallego-Hernández, Daniel (2015): The use of Corpora as translation resources: a study based on a survey of Spanish professional translators. In: Perspectives: Studies in Translatology 23 (3), 375–391.

Garcia, Ignacio (2007): Power Shifts in Web-based Translation Memory. In: Machine Translation 21 (1), 55–68.

Gornostay, Tatiana (2010): Terminology Management in Real Use. In: Proceedings of the 5th International Conference of Applied Linguistics in Science and Education, 25–26 March, St. Petersburg, Russia.

Granger, Sylviane/Paquot, Magali (eds.) (2012): Electronic Lexicography. Oxford: Oxford University Press.

Hanks, Patrick (2012): The Corpus Revolution in Lexicography. International Journal of Lexicography 25 (4), 398–436.

Héja, Eniko (2010): The Role of Parallel Corpora in Bilingual Lexicography. In Nicoletta Calzolari, Nicoletta/Choukri, Khalid/Maegaard, Bente/Mariani, Joseph/Odijk, Jan/Piperidis, Stelios/Rosner, Mike/Tapias, Daniel (eds.): Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10). Valetta, Malta: European Language Resources Association (ELRA), 2798–2805.

Mukherjee, Joybrato (2009): Anglistische Korpuslinguistik - Eine Einführung. Grundlagen der Anglistik und Amerikanistik 33. Berlin: Schmidt.

Kruger, Alet/Wallmach, Kim/Munday, Jeremy (eds.) (2011): Corpus-Based Translation Studies. London: Bloomsbury.

LeBlanc, Matthieu (2013): Translators on Translation Memory (TM). Results of an Ethnographic Study in Three Translation Services and Agencies. The International Journal of Translation and Interpreting Research 5 (2), 1–13.

Leroyer, Patrick (2015): Turning the Corpus into a Functional Component of the Dictionary: The Case of the Oenolex Wine Dictionary. In: Procedia - Social and Behavioral Sciences, 198 (24), 257–265.

Lew, Robert (2011): Online Dictionaries of English. In: Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.): e-Lexicography. The Internet, Digital Initiatives and Lexicography. London/New York: Continuum, 230–250.

Picton, Aurelie/Fontanet, Mathilde/Maradan, Mélanie/Pulitano, Donatella (2015): Corpora in Translation: addressing the Gap between the Scholars' and the Translators' Point of View. In: Corpus Use and Learning to Translate (CULT).

Prinsloo, Danie/Heid, Ulrich/Bothma, Theo/Faaß, Gertrud (2011): Interactive, dynamic electronic dictionaries for text production. In: Kosem, Iztok/Kosem, Karmen (eds.): Electronic lexicography in the 21st century: New Applications for New Users. Proceedings of eLex 2011. Bled, 10-12 November 2011, 215–220.

O'Hagan, Minako (2013): The Impact of New Technologies on Translation Studies: A technological turn? In: Millán, Carmen/Bartrina, Francesca (eds.): Routledge Handbook of Translation Studies. Oxford: Routledge. 503–518.

Olohan, Maeve (2004): Introducing corpora in translation studies. London: Routledge.

TAUS (2013): Translation Technology Landscape Report. De Rijp, The Netherlands: TAUS BV.

Temmerman, Rita/van Campenhoudt, Mark (eds.) (2014): Dynamics and Terminology. An interdisciplinary perspective on monolingual and multilingual culture-bound communication. Amsterdam/Philadelphia: Benjamin's.

Zanettin, Federico (2012): Translation-driven corpora: Corpus resources for descriptive and applied translation studies. Manchester: St. Jerome Publishing.

Zanettin, Federico/Saldanha, Gabriela/Harding, Sue-Ann (2015): Sketching landscapes in translation studies. A bibliographic study. In: Perspectives: Studies in Translatology 23 (2), 1–22.

Zaretskaya, Anna/Corpas Pastor, Gloria/Seghiri, Míriam (2016): Corpora in computer-assisted translation: a users' view. In: Corpas Pastor, Gloria/Seghiri, Míriam (eds.): Corpus-based Approaches to Translation and Interpreting. From Theory to Applications. Frankfurt am Main: Peter Lang, 253–276.

Zaretskaya, Anna/Corpas Pastor, Gloria/Seghiri, Míriam (2018): User Perspective on Translation Tools: Findings of a User Survey. In: Corpas Pastor, Gloria/Durán Muñoz, Isabel (eds.): Trends in e-tools and resources for translators and interpreters. Brill, 37–56.

Xavier Gómez Guinovart and Miguel Anxo Solla Portela

Construction of a WordNet-based multilingual lexical ontology for Galician

Abstract: This study describes the methodology used in the development of a WordNet lexicon for the Galician language, and its applications for language processing in the fields of terminology acquisition and ontology learning and management. First, we review the Princeton WordNet lexical model, its multilingual adaptation in the EuroWordNet framework, and its implementation in the Galician WordNet building. Second, we discuss the approach and the resources used in the design of Termonet, a tool for checking and verifying in technical corpora the specialty lexicons embedded in WordNet. This tool performs an identification of the synsets in WordNet belonging to a terminological domain from the semantic relations between the nodes of the lexical network, and validates the terms by means of a semantically disambiguated specialized corpus. Third, we analyze the process of construction of a new semantic categorization of WordNet based on epinonyms and generated automatically by exploring the relations from a terminological perspective. A WordNet epinonym is a noun synset in the semantic network representing the category of the semantic domain to which other synsets will be automatically assigned by algorithms that will evaluate their proximity from a terminological point of view through the cognitive processing of the lexical-semantic relations. Last, we present some applications of the RDF Galician WordNet in the Semantic Web by means of federated queries with lexical and ontological resources available as Linked Open Data (LOD) like DBpedia, BabelNet, Wiktionary and YAGO.

Keywords: computational lexicography, computational terminology, linked open data, multilingual language resources, ontologies, semantic web, WordNet

Xavier Gómez Guinovart: Grupo TALG, Universidade de Vigo, Departamento de Tradución e Lingüística, Campus Universitario, E-36310 Vigo, tel. +34 986812371, xgg@uvigo.es
Miguel Anxo Solla Portela: Grupo TALG, Universidade de Vigo, Departamento de Tradución e Lingüística, Campus Universitario, E-36310 Vigo, tel. +34 986812371, miguelsolla@uvigo.es

https://doi.org/10.1515/9783110607659-010

1 Introduction

This article[1] presents some aspects of the development of Galnet, a WordNet-based multilingual lexical ontology for Galician which is being built by the TALG Research Group (Galician Language Tecnologies and Applications) of the University of Vigo. Galnet is still under development but has already yielded interesting and useful results in the fields of Galician lexicology, semantics and automatic language processing. The following sections will describe the general lines of the project and its applications in the fields of terminology acquisition and ontology learning and management.

2 The WordNet model

WordNet is a lexical database of the English language, organized as a semantic network where the nodes are concepts represented as sets of synonyms and the links between nodes are semantic relations between lexical concepts (Fellbaum 1998, Miller et al. 1990). The nodes contain nouns, verbs, adjectives and adverbs grouped by synonymy. In WordNet terminology, a set of synonyms is called a synset, and each lemmatized synonym in a synset is called a variant and is considered a lexical variant of the same concept. Thus, each synset represents a distinct lexicalized concept and includes all the synonymous variants of this concept. Additionally, each synset may contain a brief definition or gloss, which is common to every variant in the synset, and, in some cases, one or more examples of the use of the variants in context.

In the WordNet model of lexical representation, the synsets are linked by means of lexical-semantic relations. Some of the most frequent relations represented in WordNet are hypernymy/hyponymy and holonymy/meronymy for nouns, antonymy and near synonymy for adjectives, antonymy and derivation for adverbs, and entailment, hypernymy/hyponymy, cause and opposition for verbs.

WordNet, which was originally developed for English, is now available in many other languages, although the English WordNet still stands as the most complete reference version. Created and maintained at Princeton University since 1985, version 3.0 contains 206,941 lemmas, i.e. synonymous variants (155,287 of which are unique, non-homographic forms) grouped into 117,659 sets of synonyms or synsets.

[1] This research has been carried out thanks to the project TUNER (TIN2015-65308-C5-1-R) supported by the Ministry of Economy and Competitiveness of the Spanish Government and the European Fund for Regional Development (MINECO/FEDER).

Today, WordNet is considered the most important resource in computational lexical semantics, especially in the field of natural language processing, where it is used in tasks of automatic semantic disambiguation (Agirre/Edmonds 2006), information retrieval (Zhao et al. 2012), answer extraction (Cai et al. 2016), machine translation (Vintar/Fišer/Vrščaj 2012), cross-language information retrieval (Agirre et al. 2007), cross-language question Answering (Ferrández et al. 2007), automatic text classification (Elberrichi et al. 2008), query expansion (Fang 2008), spell checking (Huang 2016), and automatic summarization (Plaza et al. 2010), among others. WordNet is also used with many benefits in computer applications related to language learning, for instance, in systems for the evaluation of the lexical competence of learners of English as a second language (Hu/Graesser 1998), for the automatic generation of gap-filling vocabulary exercises for smartphones (Knoop/Wilske 2013), or for WordNet-based vocabulary learning (Sun/Huang/Liu 2011).

Several WordNet versions are now available at different development stages in very diverse languages such as Hebrew (Ordan/Wintner 2007), Japanese (Isahara et al. 2008), Sanskrit (Bhingardive et al. 2014), Portuguese (Simões/Gómez Guinovart 2014), Spanish (Fernández/Vázquez 2010), Catalan (Oliver/Climent 2011) and Basque (Pociello/Agirre/Aldezabal 2011). Many of the WordNet versions in languages other than English follow the design model of EuroWordNet (Vossen 2002), where the synsets of a particular language are linked to the synsets of the other languages through an InterLingual Index (ILI) that is unique to each concept, and which is mainly based on the synsets of the English WordNet. Therefore, the set of WordNet lexicons in different languages allows the connection between the synsets of any pair of languages via the ILI, thus constituting a very useful resource in applications of linguistic technologies dealing with multilingual processing.

It is also worth noting that the concepts contained in the EuroWordNet database are categorized into domain hierarchies and ontologies, such as the WordNet Domains (Bentivogli et al. 2004), the Suggested Upper Merged Ontology (SUMO) (Pease/Niles/Li 2002) and the Top Concept Ontology (Álvez et al. 2008), which allows the various applications benefiting from these semantic categorizations to make better use of the resource.

3 The Galnet project

The goal of the Galnet project (Gómez Clemente et al. 2013, Solla Portela/Gómez Guinovart 2015a, Álvarez de la Granja/Gómez Clemente/Gómez Guinovart 2016) is building a WordNet for Galician aligned with the ILI generated from the English WordNet 3.0, following the *expand model* (Vossen 2002) for the creation of new wordnets, where the variants associated with the Princeton WordNet synsets are obtained from different language resources (corpora, dictionaries, glossaries…)

using experimental methods of modern lexicography developed in the field of lexical knowledge acquisition. This project is part of a wider one aimed at the coordinated integration of the Spanish, Catalan, Galician, Basque and Portuguese versions of WordNet 3.0. The research groups participating in this project are IXA (Euskal Herriko Unibertsitatea/University of the Basque Country), TALP (Universitat Politècnica de Catalunya), GRIAL (Universitat Autónoma de Barcelona, Universitat de Barcelona, Universitat de Lleida and Universitat Oberta de Catalunya), IULA (Universitat Pompeu Fabra), and SLI/TALG (Universidade de Vigo), which is responsible for building Galnet.

Galnet is part of the Multilingual Central Repository (MCR) (González/Rigau 2013), a database that currently integrates wordnets from six different languages (English, Spanish, Catalan, Galician, Basque and Portuguese) using WordNet 3.0 as ILI and where each synset is classified under the WordNet Domains hierarchy, the SUMO ontology and the Top Concept Ontology.

Tab. 1: Synsets and variants by language

	English (WordNet 3.0)		**Galician (Galnet 3.0.24)**	
	variants	*synsets*	*variants*	*synsets*
Nouns	146,312	82,115	45,040	30,039
Verbs	25,047	13,767	6,541	2,785
Adjectives	30,002	18,156	10,039	6,135
Adverbs	5,580	3,621	1,038	706
Total	206,941	117,659	62,658	39,665
Percent	100%	100%	30%	34%
		Spanish		**Portuguese**
	variants	synsets	variants	synsets
Nouns	101,027	55,227	17,149	10,047
Verbs	20,953	9,541	8,407	3,786
Adjectives	20,938	12,373	6,330	3,581
Adverbs	3,583	1,854	789	528
Total	146,501	78,995	32,675	17,942
Percent	71%	67%	16%	15%

	Catalan		Basque	
	variants	synsets	variants	synsets
Nouns	73,810	46,917	40,420	26,710
Verbs	14,619	6,349	9,469	3,442
Adjectives	11,212	6,818	148	111
Adverbs	1152	872	0	0
Total	100,793	60,956	50,037	30,263
Percent	49%	52%	24%	26%

The specific interface designed to query Galnet[2] extends the MCR functionalities by providing different types of navigation through domain hierarchies and ontologies, allowing an interactive tree-based visualization of synsets by their semantic relations, and including temporal values and sentiment scores for synsets from TempoWordNet[3], SentiWordNet 3.0[4] and ML-SentiCon[5], a new presentation of information associated with synsets in Linked Open Data format with LodLive and Virtuoso Facets, a tool specifically designed for the extraction of lexical-semantic fields (Termonet), and a new terminology-oriented semantic categorization based on epinonyms. All these issues will be discussed in depth in the sections which follow.

Table 1 shows the current development status of WordNet in the six languages integrated in the project and available via the Galnet web interface. Table 1 includes the number of synsets and variants by language and part of speech, and shows the percentage progress obtained by each lexicon which is part of MCR with respect to the extension of English WordNet 3.0, and the updated data for Galician at Galnet version 3.0.24. It should be noted that the official distribution of Galnet done through the MCR platform, while being extremely important for the dissemination and use of the resource, is just a "frozen" version of the database, and the most up-to-date data can only be accessed directly from Galnet's web interface.

Beyond its internal network of semantic senses, Galnet is fully functional for users who are looking for purely lexicographic results. Galician variants in Galnet are linked to other multimedia resources highly relevant to Galician lexicography, such as the Dicionario da Real Academia Galega[6], the Dicionario de Pronuncia da Lingua Galega[7] (phonetic and audio transcription of monolexical lemmas), the

[2] [<http://sli.uvigo.gal/galnet>; last access: January 24, 2017].

[3] [<https://tempowordnet.greyc.fr>; last access: January 24, 2017].

[4] [<http://sentiwordnet.isti.cnr.it>; last access: January 24, 2017].

[5] [<http://timm.ujaen.es/recursos/ml-senticon/>; last access: January 24, 2017].

[6] [<https://academia.gal/dicionario>; last access: November 4, 2018].

[7] [<http://ilg.usc.es/pronuncia/>; last access: November 4, 2018].

encyclopedic contents of the Wikipedia structured through the Galician BDpedia[8], or the photo gallery illustrating each synset obtained from Imagenet[9].

The latest version of Galnet, together with other important lexical and textual resources, is also available for consultation on the RILG (Integrated Language Resources for Galician) website[10].

4 Applications in terminology acquisition

WordNet was originally conceived in the context of psycholexicology, and structured by semantic relations between synsets belonging to different grammatical categories (Miller et al. 1990). It shows, therefore, many similarities with methodological aspects of terminology in terms of conceptual typology and structure. Similar semantic relations can be found in terminological repertoires such as the SNOMED Clinical Terms[11] or the Termoteca[12] terminological database (Gómez Guinovart 2012).

These similar features led us to the idea of reorienting the relations in WordNet towards strategies for exploring terminology in the lexical-semantic network. A revision of the Galnet lexical-semantic relations was then undertaken in order to examine the possibilities of building a hierarchical structure (Cabré 1992) from all the set of the WordNet synsets, acknowledging the difficulties of working with such generic synsets in terminology research, which tends to deal with more precise concepts.

Furthermore, WordNet synsets are distributed into four different grammatical categories (83,246 noun synsets, 18,156 adjective synsets, 13,885 verb synsets and 3,621 adverb synsets in WordNet 3.0) and, despite the high density of nouns, there is a significant number of synsets that cannot be considered a priori terminological concepts. A general approach would be to try to link adjectives and adverbs, whenever possible, with a terminologically relevant synset through transcategorial lexical relations; for instance, linking the concepts *surgical* and *surgically* with *surgery*. This approach shows numerous limitations, mainly because this lexical relation is sometimes missing in WordNet, and when these morphologically-based relations are codified in WordNet they occur between English-language variants rather than between concept nodes. Where this is not possible, other semantic relations such as near synonymy or antonymy would be used, despite their tendency to change the field of specialty with certain adjectives, because our strategy for exploring termi-

[8] [<http://gl.dbpedia.org/wiki/>; last access: November 4, 2018].

[9] [<http://www.image-net.org/>; last access: November 4, 2018].

[10] [<http://sli.uvigo.gal/RILG/>; last access: January 24, 2017].

[11] [<http://www.ihtsdo.org/snomed-ct/>; last access: January 24, 2017].

[12] [<http://sli.uvigo.gal/termoteca/>; last access: January 24, 2017].

nology in WordNet includes all the synsets of the lexical-semantic network regardless of their grammatical category.

To this end, it was assumed from the very beginning that it would be necessary to extend the cognitive approach to the study of the relations with an empirical verification of the results, by examining the presence of the variants in specialized text corpora from the perspective of communicative terminology.

Therefore, a methodology was designed to perform a parameterized browsing through a domain-specific lexicon given a synset representing that specific domain. Also, a method was sought to verify the empirical occurrence of the concepts in specialized Galician-language corpora. Terminology browsing from a given synset and verification in selected domain-specific corpora have been implemented in a freely accessible web application called Termonet[13], which is integrated into Galnet's public interface (Solla Portela/Gómez Guinovart 2015b).

Termonet's features rely on two basic resources: the latest development version of Galnet and the Galician Technical Corpus[14], a lemmatized, POS-tagged and semantically annotated terminology-oriented corpus of 18 million tokens containing contemporary specialized Galician texts in the fields of law, computing, economics, environmental science, sociology and medicine. The Galician Technical Corpus was tagged using Freeling[15] and UKB (Agirre/Soroa 2009) and disambiguated with respect to the Galnet lexicon.

The main function of Termonet is to enable the extraction of domain-specific variants from WordNet. For this purpose, Termonet provides a query form that allows selecting a synset from the lexical-semantic network and extracting related terms according to the semantic relations defined in the configuration. Although Termonet allows extraction from any WordNet synset, due to its terminological nature, the application always tries to suggest the closest noun variants when initiating a search from a non-noun synset.

Termonet allows to indicate the source synset that will determine the domain-specific term extraction, and also to select the semantic relations that will be used to identify the domain-specific terms and the maximum distance or depth level to be reached for each relation type. Thus, Termonet will display the tree-like structure of relations from the source synset through the selected relations to the specified depth level in the tree. Figure 1 shows, for instance, the hyponymy relation found among linguistic terms, starting at the synset represented by the English variant *linguistics* according to the system default parameters. For reasons of space, the output of the extraction illustrated in Figure 1 includes only its first results, since the total number

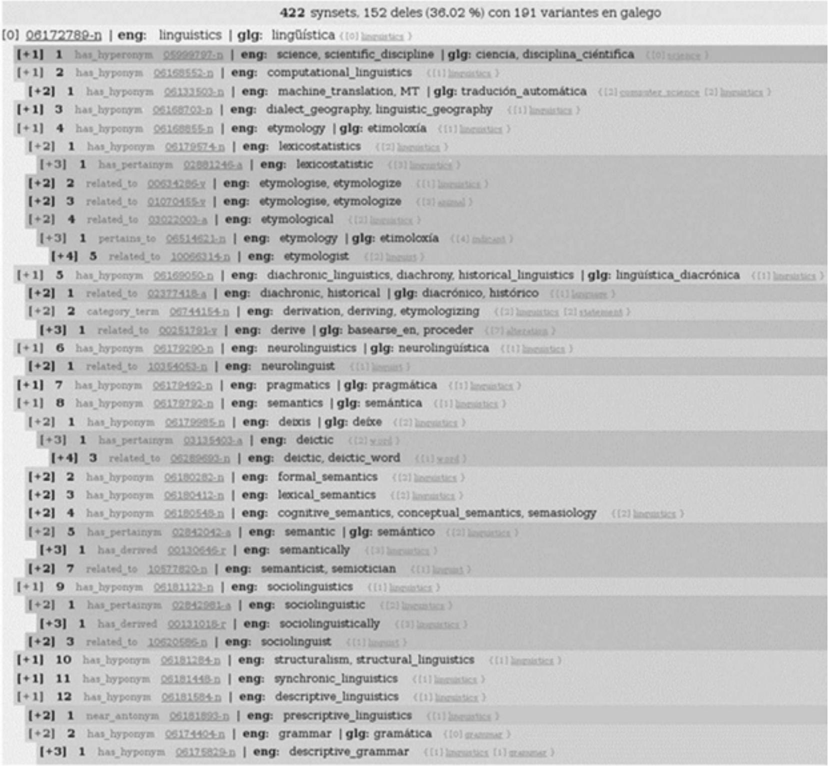
[13] [<http://sli.uvigo.gal/galnet/termonet.php>; last access: January 24, 2017].

[14] [<http://sli.uvigo.gal/CTG/>; last access: January 24, 2017].

[15] [<http://nlp.lsi.upc.edu/freeling/>; last access: January 24, 2017].

of synsets identified by this query amounts to 422[16]. Each line of Figure 1 describes a synset in the domain of *linguistics* using the following information separated by spaces:

1. distance (or depth level) from the synset described to the source synset representing the selected area of exploration;
2. lexical-semantic relation linking the synset described with the previous synset in the extraction chain;
3. ILI that identifies the synset described;
4. variants of the synset in English;
5. variants of the synset in Galician, if any;
6. semantic class or epinonym to which the synset described is linked, with the distance from the synset to the epinonym indicated in brackets.



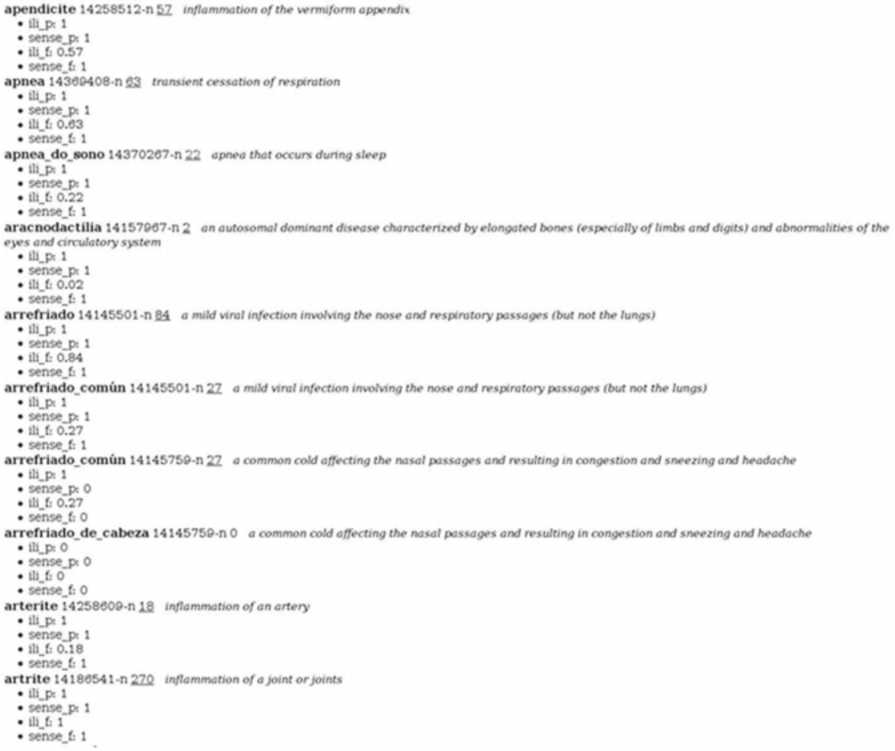
```
422 synsets. 152 deles (36.02 %) con 191 variantes en galego
[0] 06172789-n | eng:  linguistics | glg:  lingüística ([0] linguistics )
  [+1] 1 has_hyperonym 05999797-n | eng:  science, scientific_discipline | glg: ciencia, disciplina_científica ([0] science )
  [+1] 2 has_hyponym 06168552-n | eng:  computational_linguistics ([1] linguistics )
    [+2] 1 has_hyponym 06133503-n | eng:  machine_translation, MT | glg: tradución_automática ([2] computer_science [2] linguistics )
  [+1] 3 has_hyponym 06168703-n | eng:  dialect_geography, linguistic_geography ([1] linguistics )
  [+1] 4 has_hyponym 06168855-n | eng:  etymology | glg: etimoloxía ([1] linguistics )
    [+2] 1 has_hyponym 06179574-n | eng:  lexicostatistics ([2] linguistics )
      [+3] 1 has_pertainym 02881246-a | eng:  lexicostatistic ([3] linguistics )
    [+2] 2 related_to 00634286-v | eng:  etymologise, etymologize ([1] linguistics )
    [+2] 3 related_to 01070455-v | eng:  etymologise, etymologize ([3] animal )
    [+2] 4 related_to 03022003-a | eng:  etymological ([2] linguistics )
      [+3] 1 pertains_to 06514621-n | eng:  etymology | glg: etimoloxía ([4] indicant )
        [+4] 5 related_to 10066314-n | eng:  etymologist ([2] linguist )
  [+1] 5 has_hyponym 06169050-n | eng:  diachronic_linguistics, diachrony, historical_linguistics | glg: lingüística_diacrónica ([1] linguistics )
    [+2] 1 related_to 02377418-a | eng:  diachronic, historical | glg: diacrónico, histórico ([1] language )
    [+2] 2 category_term 06744154-n | eng:  derivation, deriving, etymologizing ([2] linguistics [2] statement )
      [+3] 1 related_to 00251791-v | eng:  derive | glg: basearse_en, proceder ([7] alteration )
  [+1] 6 has_hyponym 06179290-n | eng:  neurolinguistics | glg: neurolingüística ([1] linguistics )
    [+2] 1 related_to 10354053-n | eng:  neurolinguist ([1] linguist )
  [+1] 7 has_hyponym 06179492-n | eng:  pragmatics | glg: pragmática ([1] linguistics )
  [+1] 8 has_hyponym 06179792-n | eng:  semantics | glg: semántica ([1] linguistics )
    [+2] 1 has_hyponym 06179985-n | eng:  deixis | glg: deixe ([2] linguistics )
      [+3] 1 has_pertainym 03135403-a | eng:  deictic ([2] word )
        [+4] 3 related_to 06289693-n | eng:  deictic, deictic_word ([1] word )
    [+2] 2 has_hyponym 06180282-n | eng:  formal_semantics ([2] linguistics )
    [+2] 3 has_hyponym 06180412-n | eng:  lexical_semantics ([2] linguistics )
    [+2] 4 has_hyponym 06180548-n | eng:  cognitive_semantics, conceptual_semantics, semasiology ([2] linguistics )
    [+2] 5 has_pertainym 02842042-a | eng:  semantic | glg: semántico ([2] linguistics )
      [+3] 1 has_derived 00130646-r | eng:  semantically ([3] linguistics )
    [+2] 7 related_to 10577820-n | eng:  semanticist, semiotician ([1] linguist )
  [+1] 9 has_hyponym 06181123-n | eng:  sociolinguistics ([1] linguistics )
    [+2] 1 has_pertainym 02842981-a | eng:  sociolinguistic ([2] linguistics )
      [+3] 1 has_derived 00131018-r | eng:  sociolinguistically ([3] linguistics )
    [+2] 3 related_to 10620586-n | eng:  sociolinguist ([1] linguist )
  [+1] 10 has_hyponym 06181284-n | eng:  structuralism, structural_linguistics ([1] linguistics )
  [+1] 11 has_hyponym 06181448-n | eng:  synchronic_linguistics ([1] linguistics )
  [+1] 12 has_hyponym 06181584-n | eng:  descriptive_linguistics ([1] linguistics )
    [+2] 1 near_antonym 06181893-n | eng:  prescriptive_linguistics ([1] linguistics )
    [+2] 2 has_hyponym 06174404-n | eng:  grammar | glg: gramática ([0] grammar )
      [+3] 1 has_hyponym 06175829-n | eng:  descriptive_grammar ([1] linguistics [1] grammar )
```

Fig. 1: Term extraction

[16] The full results of this query are available at [<http://sli.uvigo.gal/galnet/termonet.php?ili=ili-30-06172789-n>; last access: January 24, 2017].

The application was designed to be interactive. The user can restrict the relations in WordNet in order to define his own conceptual field. After selecting the initial synset, which represents the area of the inquiry, the query form offers the different parameters of configuration for the groups of relations and for each relation in two distinct sections. First, the distance of the exploration is defined in order to limit the navigation towards conceptual nodes tied by means of ascending relations and to determine the extension of the descent and of the navigation in horizontal direction. The second section of the form allows to indicate the groups of relations or the individual relations that should be omitted from the query. It also offers the possibility of filtering the synsets with the grammatical categories selected for each relation.

This term extraction tool, albeit still in the testing phase, is already producing consistent and quantitatively significant results using rather simple search parameter settings. The extraction procedure is identical for both broad (e.g. biology) and narrow (e.g. microbiology) conceptual domains.



apendicite 14258512-n 57 *inflammation of the vermiform appendix*
- ili_p: 1
- sense_p: 1
- ili_f: 0.57
- sense_f: 1

apnea 14369408-n 63 *transient cessation of respiration*
- ili_p: 1
- sense_p: 1
- ili_f: 0.63
- sense_f: 1

apnea_do_sono 14370267-n 22 *apnea that occurs during sleep*
- ili_p: 1
- sense_p: 1
- ili_f: 0.22
- sense_f: 1

aracnodactilia 14157967-n 2 *an autosomal dominant disease characterized by elongated bones (especially of limbs and digits) and abnormalities of the eyes and circulatory system*
- ili_p: 1
- sense_p: 1
- ili_f: 0.02
- sense_f: 1

arrefriado 14145501-n 84 *a mild viral infection involving the nose and respiratory passages (but not the lungs)*
- ili_p: 1
- sense_p: 1
- ili_f: 0.84
- sense_f: 1

arrefriado_común 14145501-n 27 *a mild viral infection involving the nose and respiratory passages (but not the lungs)*
- ili_p: 1
- sense_p: 1
- ili_f: 0.27
- sense_f: 1

arrefriado_común 14145759-n 27 *a common cold affecting the nasal passages and resulting in congestion and sneezing and headache*
- ili_p: 1
- sense_p: 0
- ili_f: 0.27
- sense_f: 0

arrefriado_de_cabeza 14145759-n 0 *a common cold affecting the nasal passages and resulting in congestion and sneezing and headache*
- ili_p: 0
- sense_p: 0
- ili_f: 0
- sense_f: 0

arterite 14258609-n 18 *inflammation of an artery*
- ili_p: 1
- sense_p: 1
- ili_f: 0.18
- sense_f: 1

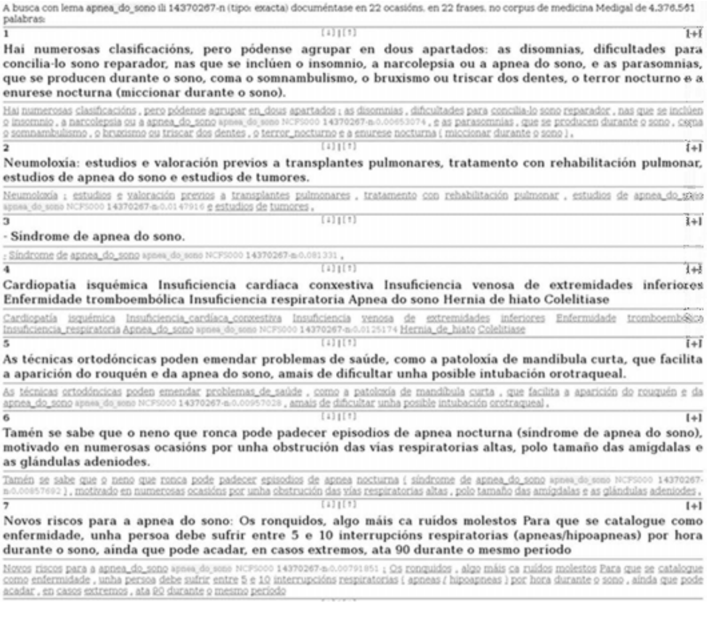
artrite 14186541-n 270 *inflammation of a joint or joints*
- ili_p: 1
- sense_p: 1
- ili_f: 1
- sense_f: 1

Fig. 2: Term evaluation

As previously mentioned, Termonet allows to verify the extraction results in a domain-specific textual corpus that has been lemmatized and disambiguated with WordNet senses. Termonet evaluates the occurrence of each monolexical or plurilexical term in the selected domain-specific corpus by applying four criteria scored from 0 to 1, and combines the results obtained across all the terms into a general score for each criterion. The four evaluation criteria are as follows:

1. *ili*-p index: the variant is present (1) or absent (0) as a lemma in the corpus and is semantically tagged with the corresponding synset.
2. *sense*-p index: the variant is present as a lemma in the corpus and is tagged with the most likely sense (1) or not (0) according to UKB.
3. *ili*-f index: absolute frequency of the variant in the corpus, awarding the maximum score (1) to the sense-tagged variants occurring at least 100 times, and the minimum score (0) to the variants not found in the corpus.
4. *sense*-f index: frequency of UKB assigning the maximum likelihood to the variant's synset tag, with the maximum score (1) for all the times and the minimum (0) for none.

Through a detailed analysis of the variants (Figure 2), Termonet offers the possibility of exploring their contexts of use in the selected domain-specific corpus of the Galician Technical Corpus (Figure 3), thus providing valuable terminological information about the real use of the terms.



A busca con lema apnea_do_sono ili 14370267-n (tipo: exacta) documéntase en 22 ocasións. en 22 frases. no corpus de medicina Medigal de 4.376.561 palabras:

1 [↓]|[↑] [+]

Hai numerosas clasificacións, pero pódense agrupar en dous apartados: as disomnias, dificultades para concilia-lo sono reparador, nas que se inclúen o insomnio, a narcolepsia ou a apnea do sono, e as parasomnias, que se producen durante o sono, coma o somnambulismo, o bruxismo ou triscar dos dentes, o terror nocturno e a enurese nocturna (miccionar durante o sono).

Hai numerosas clasificacións , pero pódense agrupar en_dous apartados : as disomnias , dificultades para concilia-lo sono reparador , nas que se inclúen o insomnio , a narcolepsia ou a apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.00653074 , e as parasomnias , que se producen durante o sono , coma o somnambulismo , o bruxismo ou triscar dos dentes , o terror_nocturno e a enurese nocturna (miccionar durante o sono) .

2 [↓]|[↑] [+]

Neumoloxía: estudios e valoración previos a transplantes pulmonares, tratamento con rehabilitación pulmonar, estudios de apnea do sono e estudios de tumores.

Neumoloxía : estudios e valoración previos a transplantes pulmonares , tratamento con rehabilitación pulmonar , estudios de apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.0147916 e estudios de tumores .

3 [↓]|[↑] [+]

- Síndrome de apnea do sono.

- Síndrome de apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.081331 .

4 [↓]|[↑] [+]

Cardiopatía isquémica Insuficiencia cardíaca conxestiva Insuficiencia venosa de extremidades inferiores Enfermidade tromboembólica Insuficiencia respiratoria Apnea do sono Hernia de hiato Colelitiase

Cardiopatía isquémica Insuficiencia_cardíaca_conxestiva Insuficiencia venosa de extremidades inferiores Enfermidade tromboembólica Insuficiencia_respiratoria Apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.0125174 Hernia_de_hiato Colelitiase

5 [↓]|[↑] [+]

As técnicas ortodóncicas poden emendar problemas de saúde, como a patoloxía de mandíbula curta, que facilita a aparición do rouquén e da apnea do sono, amais de dificultar unha posible intubación orotraqueal.

As técnicas ortodóncicas poden emendar problemas_de_saúde , como a patoloxía de mandíbula curta , que facilita a aparición do rouquén e da apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.00957028 , amais de dificultar unha posible intubación orotraqueal .

6 [↓]|[↑] [+]

Tamén se sabe que o neno que ronca pode padecer episodios de apnea nocturna (síndrome de apnea do sono), motivado en numerosas ocasións por unha obstrución das vías respiratorias altas, polo tamaño das amígdalas e as glándulas adeniodes.

Tamén se sabe que o neno que ronca pode padecer episodios de apnea nocturna (síndrome de apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.00857692) , motivado en numerosas ocasións por unha obstrución das vías respiratorias altas , polo tamaño das amígdalas e as glándulas adeniodes .

7 [↓]|[↑] [+]

Novos riscos para a apnea do sono: Os ronquidos, algo máis ca ruídos molestos Para que se catalogue como enfermidade, unha persoa debe sufrir entre 5 e 10 interrupcións respiratorias (apneas/hipoapneas) por hora durante o sono, aínda que pode acadar, en casos extremos, ata 90 durante o mesmo período

Novos riscos para a apnea_do_sono apnea_do_sono NCFS000 14370267-n 0.00791851 : Os ronquidos , algo máis ca ruídos molestos Para que se catalogue como enfermidade , unha persoa debe sufrir entre 5 e 10 interrupcións respiratorias (apneas / hipoapneas) por hora durante o sono , aínda que pode acadar , en casos extremos , ata 90 durante o mesmo período

Fig. 3: Terms in context

5 Applications in ontology learning and management

5.1 Applications in WordNet semantic areas

A new semantic categorization of WordNet was devised to exploit the terminological implications of the relations between synsets. The adopted approach was based on tracing a path in the opposite direction to that used by Termonet to explore a domain from a synset, so each synset finds its way through the relations to an *epinonym* noun synset representing the semantic domain where it is to be included. Thus, we define an *epinonym* as a noun synset representing the category of the semantic domain to which other synsets will be automatically assigned by algorithms that will evaluate their proximity from a terminological point of view through the cognitive processing of the lexical-semantic relations.

There is a certain parallelism between the method used for the selection of the semantic classes designed for Basic Level Concepts (BLC) (Izquierdo/Suárez/Rigau 2015) and for epinonyms. In both cases, there is no preset semantic categorization, and synsets are selected by the fact of fulfilling certain criteria of selection. Furthermore, in both cases, one of the selection criteria is the computation of lexical-semantic relations. However, BLC and epinonyms have a substantially different objective. BLC are nominal and verbal synsets selected for being semantically representatives, and for being neither too concrete nor too abstract. For its part, epinonyms are nominal synsets selected by its aptitude to represent a semantic area, preferably a terminologically relevant area of specialty. The selection of epinonyms is limited to the grammatical category of nouns because only nominal synsets are considered appropriate to designate areas of specialty. In addition, its coverage tries to be representative of all semantic areas, examining the entire depth of the hyponymy tree of nouns.

This exploration of the terminological implications of the relations between synsets is carried out in two different phases:

1. Automatic selection of the epinonym noun synsets to become the categories representing the different semantic domains and grouping the synsets assigned to each of the domains.
2. Development of a set of algorithms to calculate the path from each WordNet synset through the lexical-semantic relations to the closest epinonymic category or categories.

The algorithm used for selecting the set of epinonyms explores the noun synsets downward through the hyponymy relation; that is, it departs from the synset that corresponds to the concept *entity*, which is at the top (or level 0) of the hypernymy relations among WordNet noun synsets, and travels through the hyponymy relations assigning higher values of hyponymy as the distance to the top increases. The selection of the epinonyms is based on a score assigned to each synset according to the number of its semantic relations with other synsets, and to its hyponymy level. This approach yielded a hierarchical tree of 927 epinonyms, which represent 1.11% of the noun synsets in WordNet (83,246) and 0.78% of all WordNet synsets (118,868). By starting at the top level of hypernymy, the algorithm scheme preserves the inheritance in the WordNet hyponymy tree so that the relations between the selected epinonyms may be rebuilt into a hierarchical tree of subcategories, the structure and extension of which may be viewed online via Galnet's public interface[17].

The selected set of epinonyms provides an overview of WordNet semantic areas. It is consistent with its own internal structure and is generated automatically by exploring the relations from a terminological perspective; therefore, there is no modeling of the lexical-semantic network to conform to preconceived categorial notions.

After a representative set of noun synsets was obtained, a methodology was developed to link each WordNet synset to the closest epinonym(s) according to terminological criteria based on the characteristics of the lexical-semantic relations. The algorithm used to assign epinonyms generated 128,986 pairings, which is slightly above the total number of WordNet synsets (108.51%), given that some synsets were linked with more than one epinonym. Furthermore, 99.80% of the WordNet synsets were able to automatically assign themselves to their corresponding semantic area(s) and only 239 synsets (225 adverbial, 8 verbal and 6 adjectival) were not paired with an epinonym, as shown in Table 2.

Tab. 2: Pairings between synsets and epinonyms

	Paired synsets	WordNet synsets	Percent
Nouns	83,246	83,246	100%
Verbs	18,150	18,156	99.97%
Adjectives	13,837	13,845	99.94%
Adverbs	3,396	3,621	93.79%
Total	118,629	118,868	99.80%

Owing to the distinctive ontological-relational nature of the results, alongside the ontologies already present in Galnet's interface, additional search methods were

[17] [<http://sli.uvigo.gal/galnet/hierarchy.php?version=dev&ontology=epinonyms&category=entity ##ili- 30-00001740-n>; last access: January 24, 2017].

included to enable exploring the design of WordNet semantic areas represented by the epinonyms. In addition, similarly to what is described in Section 4 for Termonet, a web application[18] was built to enable the verification of Galician variants in specialized corpora by selecting all the variants of a category from any of the ontologies.

5.2 Applications in the Semantic Web

We also provide all contents of Galnet as RDF (Resource Description Framework) resources[19] through a SPARQL endpoint[20] with free public access for users to explore our data using SPARQL queries. Turtle files with data corresponding to the latest public release of MCR and related ontologies can also be downloaded from our site[21].

The RDF Galnet monolingual dictionaries conform to the Lemon (Lexicon Model for Ontologies) model[22]. The Galnet synsets are aligned with Princeton's WordNet synsets version 3.1[23], with Princeton's WordNet synsets version 3.0 in lemonUby[24] and with the Interlingual Index (ILI). In many cases, Princeton's WordNet also provides the aligment with a corresponding synset in lemonUby version 3.0. However, the alignment in RDF Galnet offers correspondences between all MCR synsets from version 3.0 and version 3.1 or lemonUby ones.

The RDF Galnet internal ontology is based on the ontology that uses the RDF Princeton Wordnet 3.1, revised and adapted to the EuroWordNet framework followed by the MCR project. Moreover, all the ontologies linked to the Galnet synsets were converted to the RDF data model: Adimen-SUMO, Top Ontology, WordNet Domains and Epinonyms.

Query 1: SPARQL federated query for Galician variant trabe

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
SELECT * WHERE {
 SERVICE <http://sli.uvigo.gal/sparql> {
```

[18] [<http://sli.uvigo.gal/galnet/category.php>; last access: January 24, 2017].
[19] [<https://www.w3.org/RDF/>; last access: January 24, 2017].
[20] [<http://sli.uvigo.gal/sparql/>; last access: January 24, 2017].
[21] [<http://sli.uvigo.gal/lod/rdf_galnet_mcr30-2016.7z>; last access: January 24, 2017].
[22] [<http://lemon-model.net>; last access: January 24, 2017].
[23] [<https://wordnet.princeton.edu>; last access: January 24, 2017].
[24] [<http://lemon-model.net/lexica/uby/>; last access: January 24, 2017].

```
  GRAPH <http://sli.uvigo.gal/rdf_galnet_glg> {
   ?offset rdfs:label "trabe"@glg .
   ?offset owl:sameAs ?pwn .
   FILTER (REGEX(str (?pwn), "^http://wordnet-rdf.princeton.edu/wn31/"))
  }
 }
 SERVICE <http://babelnet.org/sparql/> {
  GRAPH <http://babelnet.org/rdf/> {
   ?babelnet <http://babelnet.org/model/babelnet#wiktionaryPageLink>
?wiktionary .
   ?babelnet skos:exactMatch ?pwn .
   ?babelnet skos:exactMatch ?eng_resource .
  }
 }
 SERVICE <http://gl.dbpedia.org/sparql> {
  GRAPH <http://gl.dbpedia.org> {
   ?glg_resource owl:sameAs ?eng_resource .
   FILTER (REGEX(str(?glg_resource), "^http://gl.dbpedia.org/resource/")
) .
   FILTER (REGEX(str(?eng_resource), "^http://dbpedia.org/resource/") )
.
  }
 }
}
```

The availability of the data from Galnet in the semantic web allows to link the resource with other Linked Open Data and to extract a wider knowledge through SPARQL queries. For instance, Query 1 demonstrates the power of SPARQL federated queries to link a Galician variant from Galnet with the corresponding resources in Princeton's WordNet, BabelNet, English Wiktionary and Dbpedia.

On the other hand, Query 2 starts from the English variant *penicillin* and extracts its ILI from the MCR, its equivalent variant in Galician, the corresponding resource in the Galician DBpedia, the IUPAC (International Union of Pure and Applied Chemistry) designation in Galician of the chemical compound, the equivalent resource in the YAGO (Yet Another Great Ontology) knowledge base (Mahdisoltani/Biega/Suchanek 2015), and the 57 lexical forms in different languages listed in YAGO to designate the concept.

Query 2: SPARQL federated query for English variant penicillin

```
PREFIX  rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT * WHERE {
 SERVICE <http://sli.uvigo.gal/sparql> {
  GRAPH <http://sli.uvigo.gal/rdf_galnet_eng> {
   ?eng_offset rdfs:label "penicillin"@eng .
  }
  GRAPH <http://sli.uvigo.gal/rdf_galnet> {
   ?ili owl:sameAs ?eng_offset .
   ?ili owl:sameAs ?glg_offset .
   FILTER (REGEX(STR(?glg_offset),
"^http://sli.uvigo.gal/rdf_galnet_glg")) .
  }
  GRAPH <http://sli.uvigo.gal/rdf_galnet_glg> {
   ?glg_offset rdfs:label ?glg_label .
  }
 }
 BIND (IRI(CONCAT("http://gl.dbpedia.org/resource/",
UCASE(SUBSTR(STR(?glg_label), 1, 1)), SUBSTR(STR(?glg_label), 2))) AS
?glg_resource) .
 SERVICE <http://gl.dbpedia.org/sparql> {
  GRAPH <http://gl.dbpedia.org> {
   ?glg_resource dbo:iupacName ?glg_iupac .
  }
 }
 BIND(CONCAT("1", SUBSTR(STRAFTER(STR(?eng_offset),
"http://sli.uvigo.gal/rdf_galnet_eng/"), 1, 8) ) AS ?yago_id) .
 SERVICE <http://linkeddata1.calcul.u-psud.fr/sparql> {
  GRAPH <http://www.yago-knowledge.org> {
   ?yago_resource <http://yago-knowledge.org/resource/hasSynsetId> ?ya-
go_id  .
   ?yago_resource rdfs:label ?lang_label .
  }
 }
}
```

Finally, it is interesting to note that the joint exploitation of Linked Open Data can be an efficient, robust and reliable way to develop new wordnets or to expand the lexical coverage of existing ones (Solla Portela/Gómez Guinovart 2016).

6 Conclusion

Natural language processing plays an increasingly important role in the information society. The need for a computational representation of general and specialized lexical information capable of being exploited by language technologies (in automatic and computer-assisted translation applications, multilingual retrieval of information, semantic web, etc.) is a new challenge for the research in lexicology. The use of ontologies, understood as formal and shared specifications of the conceptualization of a domain that can be transmitted between people and/or systems, offers a suitable solution for this task. This is particularly relevant in a highly multilingual context such as the current information society, where the symbiosis between lexicon and ontologies allows us to work with powerful conceptual and methodological tools towards the best representation of general and specialized multilingual knowledge.

7 Bibliography

Agirre, Eneko/Edmonds, Peter (2006): *Word Sense Disambiguation*. Berlin: Springer.

Agirre, Eneko/Alegria, Iñaki/Rigau, German/Vossen, Piek (2007): MCR for CLIR. In: *Procesamiento del Lenguaje Natural* 38, 3–15.

Agirre, Eneko/Soroa, Aitor (2009): Personalizing PageRank for Word Sense Disambiguation. In: Schlangen, David/Oflazer, Kemal (eds.): *Proceedings of the 12th Conference of the European Chapter of the ACL*. Athens: ACL, 33–41.

Álvarez de la Granja, María/Gómez Clemente, Xosé María/Gómez Guinovart, Xavier (2016): Introducing Idioms in the Galician WordNet: Methods, Problems and Results. In: *Open Linguistics* 2 (1), 253–286.

Álvez, Javier/Atserias, Jordi/Carrera, Jordi/Climent, Salvador/Oliver, Antoni/Rigau, German (2008): Consistent Annotation of EuroWordNet with the Top Concept Ontology. In: Tanács, Attila/Csendes, Dóra/Vincze, Veronika/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the 4th Global WordNet Conference*. Szeged: Global WordNet Association, n.p.

Bentivogli, Luisa/Forner, Pamela/Magnini, Bernardo/Pianta, Emanuele (2004): Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In: Sérasset, Gilles/Armstrong, Susan/Boitet, Christian/Popescu-Belis, Andrei/Tufis, Dan (eds.): *Proceedings of COLING Workshop on Multilingual Linguistic Resources*. Geneva: ACL, 101–108.

Bhingardive, Sudha/Ajotikar, Tanuja/Kulkarni, Irawati/Kulkarni, Malhar/Bhattacharyya, Pushpak (2014): Semi-Automatic Extension of Sanskrit Wordnet using Bilingual Dictionary. In: Orav, Heili/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the Seventh Global WordNet Conference*. Tartu: Global WordNet Association, 324–329.

Cabré, Maria Teresa (1992): *La terminologia. La teoria, els mètodes, les aplicacions*. Barcelona: Empúries.

Cai, Qingqing/Gung, James/Guan, Maochen/Kurlandski, Gerald/Pease, Adam (2016): Word Substitution in Short Answer Extraction: A WordNet-based Approach. In: Barbu Mititelu, Verginica/Forăscu, Corina/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the Eighth Global WordNet Conference*. Bucarest: University of Iaşi, 66–73.

Elberrichi, Zakaria/Rahmoun, Abdelattif/Bentaalah, Mohamed Amine (2008): Using WordNet for Text Categorization. In: *The International Arab Journal of Information Technology* 5 (1), 16–24.

Fang, Hui (2008): A Re-examination of Query Expansion Using Lexical Resources. In: Arisoy, Ebru/Maier, Wolfgang/Inoue, Keisuke (eds.): *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics*. Columbus: Association for Computational Linguistics, 139–147.

Fellbaum, Christiane (ed.) (1998): *WordNet: An Electronic Lexical Database*. Cambridge: MIT Press.

Fernández Montraveta, Ana/Vázquez, Gloria (2010): La construcción del WordNet 3.0 en español. In: Castillo, María Auxiliadora/García Platero, Juan Manuel (eds.): *La lexicografía en su dimensión teórica*. Málaga: Universidad de Málaga, 201–220.

Ferrández, Sergio/Ferrández, Antonio/Roger, Sandra/López–Moreno, Pilar (2007): Búsqueda de respuestas bilingüe basada en ILI, el sistema BRILI. In: *Procesamiento del Lenguaje Natural* 38, 27–33.

Gómez Clemente, Xosé María/Gómez Guinovart, Xavier/González Pereira, Andrea/Taboada Lorenzo, Verónica (2013): Sinonimia e rexistros na construción do WordNet do galego. In: *Estudos de Lingüística Galega* 5, 27–42.

Gómez Guinovart, Xavier (2012): A Hybrid Corpus-Based Approach to Bilingual Terminology Extraction. In: Moskowich-Spiegel, Isabel/Crespo, Begoña (eds.): *Encoding the Past, Decoding The Future: Corpora in the 21st Century*. Newcastle upon Tyne: Cambridge Scholar Publishing, 147–175.

González Agirre, Aitor/Rigau, German (2013): Construcción de una base de conocimiento léxico multilingüe de amplia cobertura: Multilingual Central Repository. In: *Linguamática* 5 (1), 13–28.

Hu, Xiangen/Graesser, Arthur C. (1998): Using WordNet and latent semantic analysis to evaluate the conversational contributions of learners in the tutorial dialog. In: *Proceedings of the International Conference on Computers in Education*, vol. 2. Beijing: Springer, 337–341.

Huang, Bill (2016): WNSpell: a WordNet-Based Spell Corrector. In: Barbu Mititelu, Verginica/Forăscu, Corina/Fellbaum, Christiane/Vossen, Piek (eds.): *Proceedings of the Eighth Global WordNet Conference*. Bucarest: University of Iaşi, 135–142.

Isahara, Hitoshi/Bond, Francis/Uchimoto, Kiyotaka/Utiyama, Masao/Kanzaki, Kyoko (2008): Development of the Japanese WordNet. In: Calzolari, Nicoletta/Choukri, Khalid/Maegaard, Bente/Mariani, Joseph/Odjik, Jan/Piperidis, Stelios/Tapias, Daniel (eds.): *Proceedings of the Sixth International Language Resources and Evaluation*. Marrakech: ELRA, n.p.

Izquierdo, Rubén/Suárez, Armando/Rigau, German (2015): Word vs. Class-Based Word Sense Disambiguation. In: *Journal of Artificial Intelligence Research* 54, 83–122.

Knoop, Susanne/Wilske, Sabrina (2013): WordGap - Automatic generation of gap-filling vocabulary exercises for mobile learning. In: Volodina, Elena/Borin, Lars/Loftsson, Hrafn (eds.): *Proceedings of the Second Workshop on NLP for Computer-Assisted Language Learning*. Linköping: Linköpings Universitet, 39–47.

Mahdisoltani, Farzaneh/Biega, Joanna/Suchanek, Fabian M. (2015): YAGO3: A Knowledge Base from Multilingual Wikipedias. In: *Proceedings of the 7th Biennial Conference on Innovative Data Systems Research*. Asilomar: CIDR, n.p.

Miller, George A. et al. (1990): Wordnet: An on-line lexical database. In: *International Journal of Lexicography* 3 (4), 235–244.

Oliver, Antoni/Climent, Salvador (2011): Construcción de los WordNets 3.0 para castellano y catalán mediante traducción automática de corpus anotados semánticamente. In: *Procesamiento del Lenguaje Natural* 47, 293–300.

Ordan, Noam/Shuly Wintner (2007): Hebrew WordNet: a Test Case of Aligning Lexical Databases Across Languages. In: *International Journal of Translation* 19 (1), 39–58.

Pease, Adam/Niles, Ian/Li, John (2002): The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications. In: Pease, Adam (ed.): *Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web*. Edmonton: AAAI, s.p.

Plaza, Laura/Díaz, Alberto/Gervás, Pablo (2010): Automatic summarization of news using WordNet concept graphs. In: *IADIS International Journal on Computer Science and Information Systems* 5 (1), 45–57.

Pociello, Elisabete/Agirre, Eneko/Aldezabal, Izaskun (2011): Methodology and Construction of the Basque WordNet. In: *Language Resources and Evaluation* 45 (2), 121–142.

Simões, Alberto/Gómez Guinovart, Xavier (2014): Bootstrapping a Portuguese WordNet from Galician, Spanish and English wordnets. In: Navarro Mesa, Juan Luis *et al.* (eds.): *Advances in Speech and Language Technologies for Iberian Languages*. Berlin: Springer, 239–248.

Solla Portela, Miguel Anxo/Gómez Guinovart, Xavier (2015a): Galnet: o WordNet do galego. Aplicacións lexicolóxicas e terminolóxicas. In: *Revista Galega de Filoloxía* 16, 169–201.

Solla Portela, Miguel Anxo/Gómez Guinovart, Xavier (2015b): Termonet: Construcción de terminologías a partir de WordNet y corpus especializados. In: *Procesamiento del Lenguaje Natural* 55, 165–168.

Solla Portela, Miguel Anxo/Gómez Guinovart, Xavier (2016): DBpedia del gallego: recursos y aplicaciones en procesamiento del lenguaje. In: *Procesamiento del Lenguaje Natural* 57, 139–142.

Sun, Koun-Tem/Huang, Yueh-Min/Liu, Ming-Chi (2011): A WordNet-Based Near-Synonyms and Similar-Looking Word Learning System. In: *Educational Technology & Society* 14 (1), 121–134.

Vintar, Špela/Fišer, Darja/Vrščaj, Aljoša (2012): Were the clocks striking or surprising?: using WSD to improve MT performance. In: Proceedings of the Joint Workshop on Exploiting Synergies between Information Retrieval and Machine Translation (ESIRMT) and Hybrid Approaches to Machine Translation (HyTra) (EACL 2012). Stroudsburg: ACL, 87–92.

Vossen, Piek (2002): WordNet, EuroWordNet and Global WordNet. In: Revue française de linguistique appliquée 7, 27–38.

Zhao, Feng/Fang, Fei/Yan, Fengwei/Jin, Hai/Zhang, Qin (2012): Expanding approach to information retrieval using semantic similarity analysis based on WordNet and Wikipedia. In: International Journal of Software Engineering and Knowledge Engineering 22 (2), 305–322.

Ignacio M. Palacios Martínez and Mario Cal Varela

Designing and compiling a terminological and multilingual dictionary for language teaching and learning: key issues and some reflections

Abstract: This article is intended to give an overview of all the decisions made regarding the design and compilation of a multilingual terminological dictionary for language teaching and learning. Firstly, the need for a reference work of this nature is justified and then information is provided on the profile of the target readers and the sources used. The following sections are concerned with its general organization and contents with particular attention to the way the information is presented. A final section deals with new developments and future plans, such as the production of a user-friendly online version which will allow for the incorporation of users' feedback and will also include new multilingual glossaries of the entries in other languages apart from those already implemented. The lists of language-related online resources will also be updated and enriched regularly to further enhance the usefulness of the dictionary.

Keywords: language learning, language teaching, terminological dictionary, multilingual dictionary

1 Introduction

The last two decades have witnessed the publication of a growing number of terminological dictionaries and glossaries of different types in different fields. By a terminological dictionary we mean, following Lehman (1996: 215), “a dictionary whose entries are constituted by the elements of a terminology”. Every lemma in such a

Ignacio M. Palacios Martínez: Universidade de Santiago de Compostela, Castelao, s/n, 15782, Santiago de Compostela (Spain), tel. +34 881811890, fax +34 881811818, ignacio.palacios@usc.es
Mario Cal Varela: Universidade de Santiago de Compostela, Castelao, s/n, 15782, Santiago de Compostela (Spain), tel. +34 881811890, fax +34 881811818, mario.cal@usc.es

https://doi.org/10.1515/9783110607659-011

dictionary is a specific term related to a particular subject or field[1]. Thus we find monolingual and bilingual dictionaries in the area of Law, Finance, Architecture, Politics, Medicine, Biology, Philosophy, Pharmacy, Chemistry, to mention just a few. It might even be the case that there currently exist as many terminological dictionaries as there are fields or subfields of knowledge. Arnts and Ficht (1989: 221) make a general distinction between specialised dictionaries that are concerned with a wide field of knowledge, for example, those dealing with technology, and those that are focused on a more limited area, such as digital electronics. It is also possible to make further classifications, based on the number of languages that are present in these works. Our dictionary would thus be considered a specialised terminological dictionary dealing with the educational field and containing a multilingual glossary, since equivalent terms in several languages are provided for all entries.

Terminology, according to Cabré (1999: 19-20), is the field concerned with the terms used in a particular area of study. These lexical items differ from common words mainly in their pragmatics, this involving differences relating to their users, contexts, and discourse types. Thus, users of common words are the general speakers of a language, whereas the users of the terms of a particular branch of knowledge are the specialists who deal with that discipline. Similarly, while common words can be used in a wide array of communicative situations and text types, the terms of a specific discipline are restricted to this and to its specialised discourse (Cabré 1999: 26).

The idea underlying these lexicographic materials is that there is a specialised jargon related to the discipline that requires careful definition and explanation to help practitioners develop a working familiarity with the area. Thus, Alcaraz/Martínez Linares (1997: i), themselves very experienced specialists in the compilation of such dictionaries[2], claim that “la terminología constituye con frecuencia el principal obstáculo para la comprensión de lo que se estudia o se analiza”, that is, the terminology of a particular discipline very often becomes the main barrier to the

[1] Lehman categorises dictionaries according to the speech community they represent and to the functions they fulfil. Within the first group he distinguishes general, dialect, sociolect (e.g. colloquial language, adolescents’ language), individual (e.g. a dictionary of Shakespeare or Cervantes) and technical dictionaries. In a second category he classifies terminological dictionaries together with dictionaries of proper names, and etymological, valency, frequency, onomasiological and bilingual dictionaries.

[2] Enrique Alcaraz was the main author or director of a wide range of specialised bilingual dictionaries. The following are some of the most important: Spanish-English dictionary of marketing publicity and mass media (Alcaraz/Hughes/Campos 1999), Spanish-English dictionary of tourism and leisure (Alcaraz et al. 2000), Spanish-English dictionary of terms of footwear and allied industries (Alcaraz et al. 2006), and Spanish-English dictionary of legal terms (Alcaraz/Hughes 2007).

understanding of what is being studied or analysed. These difficulties in understanding specialized content are clear evidence for the existence of “professional” and “academic” languages, as Alcaraz/Martínez/Yus (2007: 7-8) refer to them, which are characterised by six main defining features: specific lexis, idiosyncratic syntactic and stylistic tendencies —which may vary across languages, such as with legal language, whose syntactic structures tend to be quite complex as compared to those more typical of technical jargon (Martínez Linares 2007), particular discourse preferences (expositive, descriptive, persuasive), special communicative strategies and techniques, specific genres or text types (e.g. the contract or the sentence in legal language, the leaflet or flyer in the language of tourism, the financial statement in the world of finance) and a distinct cultural framework, the latter being determined by the epistemological community to which the specific language type belongs, and which may also vary substantially across languages. Of these six defining features, the lexical component is the most significant, in that it constitutes the core of any specialised language.

An urgent need was felt for a reference work of this type in Spanish covering the field of language teaching and learning, where we have been relying for too long on dictionaries and reference works in English or translations of such works into Spanish (e.g. Richards/Platt/Platt 1993) which often ignore key notions relevant from a local perspective while they include information that is not fully contextualised or is concerned with language teaching and learning situations that are partially or wholly alien to a Spanish readership. It was thus necessary to approach all issues of language education from our own perspective, yet without neglecting the kind of global outlook which might be applicable and adaptable to any educational setting.

In this paper we describe and account for the decisions made and the steps taken throughout the process of compilating a terminological repertoire for this rather widely defined field. We will deal in turn with issues related to the scope of the book as conceived initially and the envisioned target users, the main sources of data used, the procedure followed in the selection of terms for inclusion as main entries, and finally the overall organization of the compilation and the internal structure of individual entries.

2 The starting point

The initial goal was a manageable compilation of around a thousand key terms specifically related to the teaching and learning of languages in general. The core content should therefore consist of highly specialized terminology in this broadly defined area. However, it soon became apparent that the scope needed to be widened further to include terms from fields such as sociolinguistics or psycholinguistics, as well as terms related to statistics and academic research in general. Some common words such as *edad*, *juego* and *entrevista* have also been included in the dictionary by virtue of their direct relevance to language teaching and learning in one way or another. The final list in the first printed edition of the work thus grew to about 1,200 entries. This was intended to be an encyclopedic dictionary focusing on the subject matter, rather than a linguistic dictionary concerned with grammatical aspects of the expressions included (Arnts/Ficht 1989: 221). As a tool to help readers cope with highly specialised texts, its practical orientation should be a defining feature of the final product.

Another key criterion taken into consideration was that this terminological dictionary ought to have applicability in the teaching of any language or at least of those foreign languages that are most widely taught in Spain (English, French and German)[3]. The current state of language learning should be borne in mind, then, as well as changes taking place in the Spanish educational system. For example, since the 1990s six education laws have been passed by the Spanish Parliament, including the LOGSE (*Ley de Ordenación General del sistema educativo*, the Law of the general organization of the educational system) and the current LOMCE (*Ley Orgánica para la mejora del sistema educativo*, the Organic Law for the improvement of the education system), which is in force but still not completely implemented. Clearly, education in Spain is in a continuous state of evolution and change. As a result, not only different points of view but also new sets of terms associated to different theories, tendencies and models are constantly emerging. Apart from this, there was also a need for a bank of expressions related to language teaching/learning in different foreign languages. It was important for Spanish language education professionals to know how a specific concept is denoted in a particular foreign language so that this could be easily identified and understood when dealing with specialised texts; multilingual glossaries of all entries, included at the end of the dictionary, are intended to serve this purpose.

In sum, in terms of scope and content, the decisions made on this terminological dictionary were based on the following criteria: (i) the entries included in the dictionary should be of application not only to English but also to the most widely

[3] According to recent Eurostat data, demand for languages other than English is still rather low in Spain. In 2014 only 26% of Spanish students were learning two or more foreign languages, as compared to the average EU28 rate of 51%. In upper-secondary education, that is, the so-called Bachillerato, 97.5% of the students were studying English while only 24.2% had chosen French, and a meagre 1.7% were studying German. In some schools, other foreign languages such as Portuguese, Italian, Russian and Chinese are also offered, but the number of students is typically very low. (Source: Eurostat and UNESCO Institute for Statistics UIS, OECD. See: <http://ec.europa.eu/eurostat/statisticsexplained/index.php/Foreign_language_learning_statistics>; last access: June 24, 2016).

foreign languages studied in Spain; (ii) as an educational dictionary, regular updates should be contemplated to keep up with all the latest changes in this field in Spain derived from the promulgation of the different educational laws and regulations; (iii) the dictionary should include a multilingual component to cater for the needs of language professionals who work on a range of foreign languages for which multilingual terminological resources are currently lacking in Spain; and, lastly, (iv) given its encyclopedic and pedagogical orientation, the dictionary should not only contain clear definitions of the different entries listed in it but it should also provide illustrative examples and practical information wherever convenient. The range of target readers envisaged has, in fact, worked as an overarching principle throughout the whole process.

3 To whom is this dictionary addressed?

In addition to accuracy and comprehensiveness, the quality of terminological resources rests heavily on how well they cater for the particular needs of their potential users. The experience of the compilers as language teachers and researchers in the field of Applied Linguistics has certainly played a key role in the design of this reference work to suit the needs and expectations of real users. Five main professional profiles were specifically taken into consideration:

(i) Without a doubt a major category of users consists of L1 and L2 language teachers, as well as students at all educational levels: primary, secondary, university (Modern Languages and Literatures, Education, Psychology, Translation, Humanities), EOIs[4] (Official Language Schools) and general language schools. It is important to note that although the emphasis is placed on issues related to second language learning, many of the entries included in the dictionary are also relevant to L1 acquisition.

(ii) Scholars interested in carrying out research on first and second language acquisition will find a comprehensive coverage of research-related notions from the field. There are, for example, a considerable number of entries concerned with basic research and elementary statistics applied to the teaching of foreign languages.

[4] Escuelas Oficiales de Idiomas or Official Languages Schools are state institutions where students can learn one or several modern languages for a small fee. This network of language institutions is unique to Spain within the European community; EOI centres are found in most large Spanish towns and cities. Further information about these institutions can be obtained by visiting this website: [<http://www.educaweb.com/contenidos/educativos/aprender-idiomas/escuelas-oficiales-idiomas-territorio/>; last access: November 7, 2016].

(iii) Theoretical and applied linguists are also envisaged as likely users. Key notions related to the major schools of thought in linguistics are included to the extent that they provide the theoretical underpinnings of major approaches to language pedagogy.

(iv) Many entries included in the dictionary will prove particularly useful for materials and curriculum developers, since for these specialists it is important to be acquainted with questions such as the organization and content of the language curriculum, the teacher training programmes for pre- and in-service teachers, the progress made by new technologies in language teaching, etc.

(v) Translation specialists and practitioners may find the multilingual component of the dictionary of practical use. As noted above, this dictionary includes a multilingual dictionary of all the entries included in it. At the moment, in addition to terms in Spanish, it offers translations into 5 other languages, Galician, French, Italian, English and German. We are currently working on Polish and Russian translations, and there are plans to add Portuguese, Catalan and Basque to the list.

(vi) Finally, terminologists and scholars looking at the coining of new terms will find this dictionary an interesting resource, since on many occasions we had to coin words that were non-existent in a particular language. That was particularly the case, for example, in Galician, since due to its history and status as a minority language it is more limited lexically in a number of areas and it was not always possible to find an existing term for a particular notion. To overcome this shortcoming, we resorted at times to Portuguese, a language from the same family as Galician, and which possesses a larger number of speakers and has been wholly standardised, and is permanently in use by a large speech community in all sectors and fields. That was the case, for example, for the entry "Second Language for Specific Purposes" which was translated into Galician as "Segunda Lingua para fins específicos". As a lexicographic project, the dictionary will certainly invite reflection from terminology specialists on term variants and definitions. It may also be used as a convenient source of materials on the design and compilation of new dictionaries and glossaries.

On the whole, practical considerations have been a deciding factor when choosing what to include and how much detail to provide. In particular, entries related to teaching methods and notions with immediate applicability in teaching and learning situations tend to be enriched with a profusion of examples and detailed advice.

4 Sources used

As noted above, the initial plan for the dictionary focused on specialized terminology from the rather widely defined field of language teaching and learning. Our expertise in this area allowed for a top-bottom procedure which began with a few core networks of concepts, to which successive layers of related notions were added. The core sets are essentially those included in section 6 below as "teaching practice" and theoretical foundations of "language acquisition/learning". Additional sets were then added on research methodology, institutions, general linguistics and relevant interdisciplinary areas. Key publications were systematically explored for each of the subcomponents (see section 6 below for a complete map of content), including the following:

a. Dictionaries and encyclopedias of Applied Linguistics and Language Teaching, in both their English and Spanish versions (Richards/Platt/Platt 1993, 1997, Johnson/Johnson 1998).
b. Dictionaries of Linguistics and linguistic terms (Crystal 1980, 2010, Lewandowski 1995, Trask 1993, Alcaraz/Martínez 1997).
c. Encyclopedias and handbooks of General Linguistics (Crystal 1992, 1997, Renkema 1993, Asher/Simpson 1994, Spolsky 1999).
d. Specific dictionaries and handbooks dealing with language disorders and speech pathologies (Morris 1988, Nicolosi/Harryman/Kresheck 1978, Peña Casanova 1988).
e. Major second language teaching and acquisition handbooks (Krashen 1981, Willis 1981, Howatt 1984, Nunan 1988, 1989, 1992; Ellis 1994, Ur 1996, Cenoz/Jessner 2000).
f. Major language teaching and learning handbooks written by Spanish authors or authors based in Spain (Estaire 1990, Celaya 1992, Zanón 1993, 1999, Ribé/Vidal 1995, McLaren/Madrid 1996, Salaberri 1999, Muñoz 2000).
g. Manuals dealing with the field of statistics and its application to second language research (Butler 1985, Woods/Fletcher/Hughes 1986, Brown 1988, Bisquerra 1989, Nortes 1991).
h. Manuscripts, records and reports published by institutions directly related to the teaching and learning of languages (British Council, Goethe-Institut, Alliance Française, Instituto Cervantes, Real Academia Española, Instituto da Lingua Galega, Real Academia Galega).
i. Official documents connected with the teaching and learning of languages published by the local, national and European education authorities (Ministry of Education, Spanish autonomous governments, Council of Europe).

Careful examination of the lists of terms, the subject indexes and the tables of contents contained in these sources determined not only which concepts were to be included but also which forms for each term to record. Authoritative handbooks and monographs written by native speakers of Peninsular Spanish were particularly useful as sources of information for this purpose. Other types of publications not listed here, such as research articles and conference proceedings, for example, were not used at this stage, since scholars incorporate terminology gradually as they develop their technical competence in their field, and their spontaneous use of terms in such reporting of their work may not be a reliable indicator of term currency (Cabré 1999: 218).

Loanwords were systematically avoided even though they may occasionally occur in the specialized literature in Spanish. Thus, in spite of the sporadic use of *code-switching* or *matched guise technique* in texts in Spanish, only *alternancia* (or *cambio*) *de códigos* and *test de pares falsos* were respectively recorded in the dictionary. Geographically restricted variants, such as *prueba de apareamiento disfrafazado* (used by a well-known Chilean sociolinguist for the latter of these two concepts) were not recorded either. Otherwise, widely used variants were indicated at the beginning of each entry (e.g. *pidgin* together with *sabir*, which is incidentally the preferred term in Galician). The lists of equivalents in other languages, however, do not include variant forms, as this would have required a similarly systematic exploration of monolingual sources in each of those languages, which is well beyond the scope of the current project.

5 General structure and organization of the dictionary

In the general introduction to the first printed edition of the work (pages 7-9), Professor Enrique Alcaraz, deals with new advances in the lexicographic field and highlights some of the most noteworthy features of the dictionary: the wide range of areas discussed, its general cohesion, clarity in the definition of entries, a notable spirit of pedagogical applicability, and a well-documented framework. Figure 1 below summarizes the range of topic areas covered and also gives some idea of how the dictionary's contents are structured. At a first stage we considered the possibility of presenting the different terms grouped according to subfields or typical areas within the language teaching discipline. However, we finally decided to organize the different entries alphabetically, since we found this to be of greater practical use for the target readers. The incorporation of new entries would also be easier in this way. Nevertheless, as designers of the dictionary we kept in mind a conceptual map of the whole work[5], and to a certain extent this original concept can still be found in the rich array of cross-references and in the final list of related terms included for most entries. These elements make it possible to place concepts within the network of related notions which somehow complement and delimit their meaning(s).

For a standard or typical entry, the following information is provided:

(i) Explanation of the meaning. At times this also includes the definition of different acceptations of the term (see footnote 5). Thus, for example, in the case of the entry contexto, that is, context, three different senses of the term are provided: one referring to the context of a particular linguistic unit which is formed by the elements that precede or follow it in an utterance; a second one related to the communicative or situational context that acts as a general framework for an interaction between the participants in a communicative act; and a third related to the educational context.

(ii) Cross-references to closely related items. Thus, for instance, when we define the concept of conversation, we necessarily allude to the notion of conversation analysis. Explanations of these related items are also included in the dictionary. A great effort has been made to exploit cross-references to the limit to enhance this view of terms as nodes within systems of interrelated concepts.

(iii) Bibliographical references. For most of the entries, specific bibliographical references are provided so that users who want to explore further or expand their knowledge of the topic can consult these sources. As far as possible, we have confined our attention to major or basic references and have tried to include works produced by Spanish scholars.

In addition to the explanation of entries, which are central to any dictionary of this kind, multilingual *glossaries* for each of the entries are also included. Thus, in a final section of the printed edition, we find the equivalents of all terms in 6 different languages: Spanish, Galician, English, French, Italian and German. A team of specialists worked on this task, including teachers of English in collaboration with academics and native speakers of the languages in question. This proved to be far from a simple undertaking, in that on many occasions it was necessary to overcome terminological problems by consulting the specialised bibliography in each language, and in quite a few cases solutions had to be found when an equivalent translation or a particular notion could not be found or was known to be non-existent.

[5] Arnt and Ficht (1989: 223) refer quite extensively to the differences between dictionaries organised according to basic concepts and those arranged alphabetically; whereas in the former it is the concepts as such that play the most important role, in the latter the meaning of a term constitutes the central part of an entry and it is often the case that a single term may be equivalent to two or more different concepts. Furthermore, if in the alphabetical dictionaries the entries are presented as "lexicographic units" in the case of a dictionary that has been conceptually organised, it would be more accurate to speak of "conceptual units".

This multilingual glossary is thus intended to fill an identifiable gap in available resources, since to our knowledge such a bank of information does not currently exist, despite the fact that many scholars and teachers are very much in need of one for their work.

As an additional, innovative feature of the dictionary, and also as an intermediate stage between the printed and the online version currently in preparation, the book was completed with *6 appendices* containing a selection of Internet links and websites related to areas of interest and useful content for each of the languages present in the dictionary. These appendices are organised in 5 different sections: a) main teachers' associations and other relevant institutions or organisations; b) teacher training courses, seminars and master's programmes; c) electronic journals, forums and general distribution lists related to the teaching and learning of these languages; d) teaching and learning materials (tests, reference and pedagogical grammars, workbooks, readers, dictionaries, publishing houses, electronic libraries, etc); and, finally, e) information on courses offered for each of these languages.

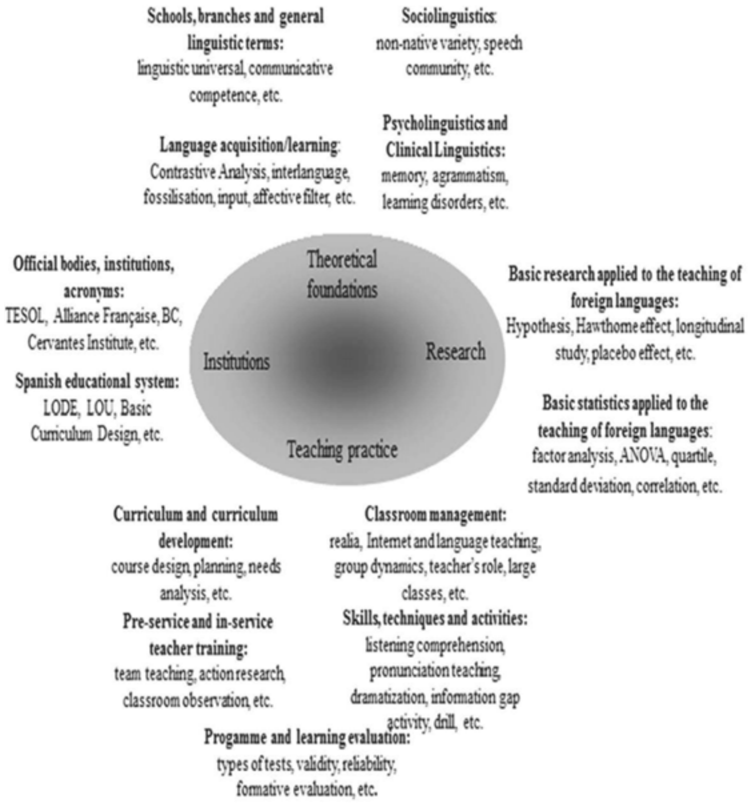

Schools, branches and general linguistic terms: linguistic universal, communicative competence, etc.
Sociolinguistics: non-native variety, speech community, etc.
Language acquisition/learning: Contrastive Analysis, interlanguage, fossilisation, input, affective filter, etc.
Psycholinguistics and Clinical Linguistics: memory, agrammatism, learning disorders, etc.
Official bodies, institutions, acronyms: TESOL, Alliance Française, BC, Cervantes Institute, etc.
Spanish educational system: LODE, LOU, Basic Curriculum Design, etc.
Theoretical foundations
Institutions
Research
Teaching practice
Basic research applied to the teaching of foreign languages: Hypothesis, Hawthorne effect, longitudinal study, placebo effect, etc.
Basic statistics applied to the teaching of foreign languages: factor analysis, ANOVA, quartile, standard deviation, correlation, etc.
Curriculum and curriculum development: course design, planning, needs analysis, etc.
Classroom management: realia, Internet and language teaching, group dynamics, teacher's role, large classes, etc.
Pre-service and in-service teacher training: team teaching, action research, classroom observation, etc.
Skills, techniques and activities: listening comprehension, pronunciation teaching, dramatization, information gap activity, drill, etc.
Progamme and learning evaluation: types of tests, validity, reliability, formative evaluation, etc.

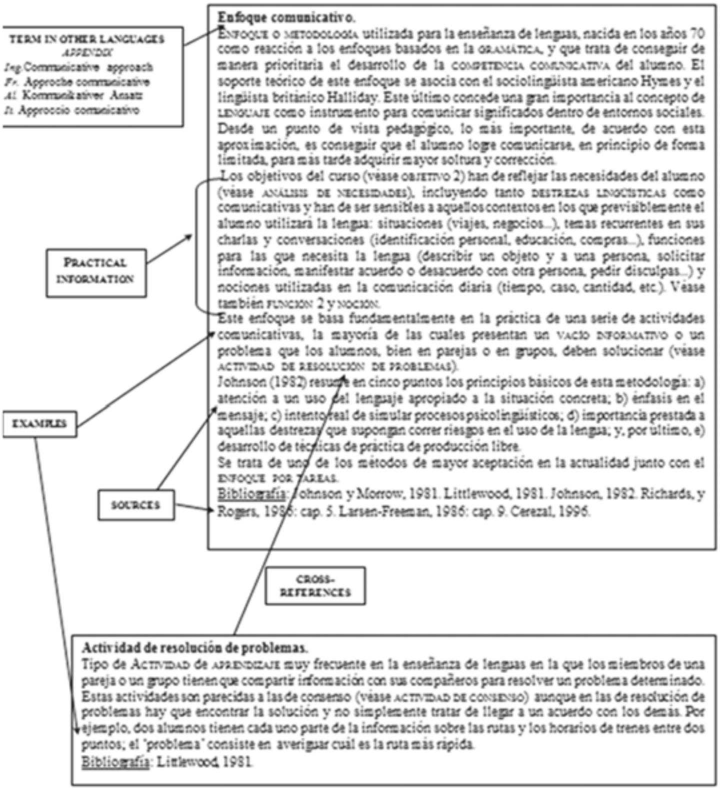
Fig. 1: Main areas and disciplines covered in the dictionary

6 General contents

As indicated above, this dictionary was conceived primarily as a reference work of an encyclopedic nature. The printed versions contain around 1,200 entries arranged in alphabetical order that provide a comprehensive list of the basic vocabulary associated with the teaching and learning of languages. It covers a wide variety of fields and subfields related to the design and evaluation of language programmes and courses, language teaching methods and techniques, teacher training, language learning factors and theories, and linguistic concepts considered to be essential for a language teaching professional. Although the simple alphabetical organization of entries was finally chosen, interrelations are constantly brought to the surface and highlighted through a dense system of cross-references as has already been pointed out. This means that in a way the dictionary might be said to collect in a single resource a set of small glossaries of research terms in linguistics, psycholinguistics, statistics, curriculum studies and language research, all of these seen from the perspective of the kind of concepts and areas of interest that may be necessary and useful for academics, scholars and language teachers in their daily activities. Figure 1 illustrates and summarizes the areas and disciplines covered.

7 How is the information presented?

In line with standard practices, entries are headed by the most usual form of the term in bold. When alternative variants are commonly used, they are given in the initial line of the body of the entry. In the printed version, equivalents for each of the entries in 6 languages are only provided in a final section of the dictionary, in tabular format. The body of each entry contains a detailed explanation of the headword. This description is complemented with practical information and examples in keeping with the pedagogical approach adopted throughout the dictionary. Finally, some bibliographical references are listed for those interested in obtaining further information. In the description of the entries, cross-references to other, directly connected concepts are also included. These are marked in small capitals. The following can be considered as an example.


TERM IN OTHER LANGUAGES
APPENDIX
PRACTICAL INFORMATION
EXAMPLES
SOURCES
CROSS-REFERENCES

Ing. Communicative approach
Fr. Approche communicative
Al. Kommunikativer Ansatz
It. Approccio comunicativo

Enfoque comunicativo.
Enfoque o metodología utilizada para la enseñanza de lenguas, nacida en los años 70 como reacción a los enfoques basados en la gramática, y que trata de conseguir de manera prioritaria el desarrollo de la competencia comunicativa del alumno. El soporte teórico de este enfoque se asocia con el sociolingüista americano Hymes y el lingüista británico Halliday. Este último concede una gran importancia al concepto de lenguaje como instrumento para comunicar significados dentro de entornos sociales. Desde un punto de vista pedagógico, lo más importante, de acuerdo con esta aproximación, es conseguir que el alumno logre comunicarse, en principio de forma limitada, para más tarde adquirir mayor soltura y corrección.

Los objetivos del curso (véase objetivo 2) han de reflejar las necesidades del alumno (véase análisis de necesidades), incluyendo tanto destrezas lingüísticas como comunicativas y han de ser sensibles a aquellos contextos en los que previsiblemente el alumno utilizará la lengua: situaciones (viajes, negocios...), temas recurrentes en sus charlas y conversaciones (identificación personal, educación, compras...), funciones para las que necesita la lengua (describir un objeto y a una persona, solicitar información, manifestar acuerdo o desacuerdo con otra persona, pedir disculpas...) y nociones utilizadas en la comunicación diaria (tiempo, caso, cantidad, etc.). Véase también función 2 y noción.

Este enfoque se basa fundamentalmente en la práctica de una serie de actividades comunicativas, la mayoría de las cuales presentan un vacío informativo o un problema que los alumnos, bien en parejas o en grupos, deben solucionar (véase actividad de resolución de problemas).

Johnson (1982) resume en cinco puntos los principios básicos de esta metodología: a) atención a un uso del lenguaje apropiado a la situación concreta; b) énfasis en el mensaje; c) intento real de simular procesos psicolingüísticos; d) importancia prestada a aquellas destrezas que supongan correr riesgos en el uso de la lengua; y, por último, e) desarrollo de técnicas de práctica de producción libre.

Se trata de uno de los métodos de mayor aceptación en la actualidad junto con el enfoque por tareas.

Bibliografía: Johnson y Morrow, 1981. Littlewood, 1981. Johnson, 1982. Richards, y Rogers, 1986: cap. 5. Larsen-Freeman, 1986: cap. 9. Cerezal, 1996.

Actividad de resolución de problemas.
Tipo de Actividad de aprendizaje muy frecuente en la enseñanza de lenguas en la que los miembros de una pareja o un grupo tienen que compartir información con sus compañeros para resolver un problema determinado. Estas actividades son parecidas a las de consenso (véase actividad de consenso) aunque en las de resolución de problemas hay que encontrar la solución y no simplemente tratar de llegar a un acuerdo con los demás. Por ejemplo, dos alumnos tienen cada uno parte de la información sobre las rutas y los horarios de trenes entre dos puntos; el "problema" consiste en averiguar cuál es la ruta más rápida.

Bibliografía: Littlewood, 1981

Fig. 2: General information on how to use the dictionary

8 Future projects and developments

The first edition of this dictionary was published in 2007 by CLE International, and was followed by a second version in 2008, sponsored by the University of Santiago de Compostela, which was not only a simple translation of the whole work into Galician but also provided an opportunity to bring the book closer to the Galician reality, by incorporating new entries directly concerned with Galician and its teaching. These new entries included, to mention just a few: ILGA (*Instituto da Lingua Galega*, 'Institute of the Galician Language') [<https://ilg.usc.es>][6]; CELGA (*Certificado de Lingua Galega*, 'Certificate of Galician') [<https://www.lingua.gal>][7]; CORGA (*Corpus*

[6] Last Access: February 11, 2017.

[7] Last Access: February 11, 2017.

de Referencia do Galego, 'Reference Corpus for Galician') [<https://corpus.cirp.es>][8]; *Atlas Lingüístico Galego* ('GalicianLinguistic Atlas'), [<https://ilg.usc.es/en/proxectos/atlas-linguistico-galego-alga>][9]; *Seccións Bilingües* ('Bilingual Sections') [<https://www.edu.xunta.gal/portal/linguasestranxeiras/>][10]; P.A.L.E. (*Plande Actualizaciónda Lingua Estranxeira*, 'Programme for the implementation of foreign languages') [<http://www.csi-f.es>][11]; VOLGa (*Vocabulario Ortográfico da Lingua Galega*, 'Spelling vocabulary of Galician') [<https://academia.gal/recursos-volg>][12], etc. In this version we also included particular examples applied specifically to Galician together with some entries concerning educational regulations and norms that are in force only in the Galician territory.

Ten years have passed by since these printed versions came out. The impact and the effects of new technologies in many areas of work in the Humanities are undeniable and the compiling of dictionaries is no exception. The trend today is towards lexicographic sources which are readily accessible on our PC or are simply available online as these electronic versions make the whole searching process quicker and more direct, so that we can continue working while we check the meaning of a word. Online versions are also more easily accessible to a much larger number of users, thus becoming more efficient tools for the consolidation of terminology. Águila Escobar (2006: 12) refers to the advantages of electronic dictionaries in terms of their flexibility, dynamism and fast and easy access, especially if compared with the rigidity of the printed versions. Moreover, an electronic dictionary may contain many other dictionaries within it and be provided with a simple interface that allows different types of searches. It is even possible to introduce multimodal and multimedia components containing not only texts but also images and sounds.

With this in mind, we are now working on a new, fully online version which will be freely available to all users. This will also give us the opportunity to implement some new features and update the content of the previous versions in order to bring it into line with almost a decade of new developments in language teaching techniques and methods (the number of entries has already increased to about 1,400). This new version will be more user-friendly, will allow users to select directly the Spanish or Galician version and browse the entries and use hyperlinks to exploit cross-references between entries much more effectively. Furthermore, readers will be given the opportunity to contact the research team and make suggestions and comments on any aspect of the content. Figure 3 is a screenshot of a sample entry showing the different components: simplified navigation menu including links to

[8] Last Access: February 11, 2017.

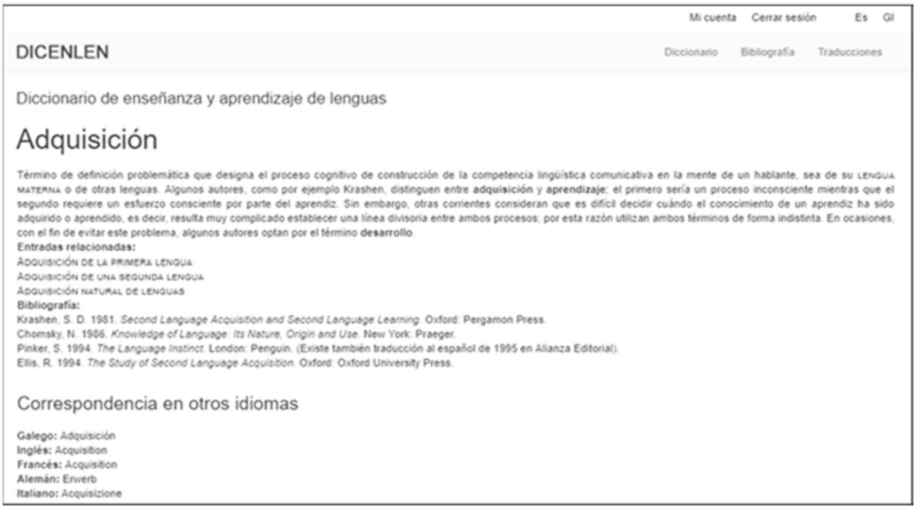
[9] Last Access: February 11, 2017.

[10] Last Access: February 11, 2017

[11] Last Access: February 11, 2017.

[12] Last Access: February 11, 2017.

toggle between the Spanish and the Galician versions (top right), entry definition (with hyperlinks in small capitals to terms with their own definition in the dictionary), list of related entries, selected bibliographical references and equivalents in other languages.



Diccionario de enseñanza y aprendizaje de lenguas

Adquisición

Término de definición problemática que designa el proceso cognitivo de construcción de la competencia lingüística comunicativa en la mente de un hablante, sea de su LENGUA MATERNA o de otras lenguas. Algunos autores, como por ejemplo Krashen, distinguen entre **adquisición** y **aprendizaje**; el primero sería un proceso inconsciente mientras que el segundo requiere un esfuerzo consciente por parte del aprendiz. Sin embargo, otras corrientes consideran que es difícil decidir cuándo el conocimiento de un aprendiz ha sido adquirido o aprendido, es decir, resulta muy complicado establecer una línea divisoria entre ambos procesos; por esta razón utilizan ambos términos de forma indistinta. En ocasiones, con el fin de evitar este problema, algunos autores optan por el término **desarrollo**.

Entradas relacionadas:
ADQUISICIÓN DE LA PRIMERA LENGUA
ADQUISICIÓN DE UNA SEGUNDA LENGUA
ADQUISICIÓN NATURAL DE LENGUAS

Bibliografía:
Krashen, S. D. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon Press.
Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger.
Pinker, S. 1994. *The Language Instinct*. London: Penguin. (Existe también traducción al español de 1995 en Alianza Editorial).
Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

Correspondencia en otros idiomas

Galego: Adquisición
Inglés: Acquisition
Francés: Acquisition
Alemán: Erwerb
Italiano: Acquisizione

Fig. 3: Sample entry of the online version under construction

This is no doubt a challenging task, but we believe that all this work will come to fruition in the near future, with the dictionary reaching a larger audience and allowing more people to take advantage of it.

9 Bibliography

9.1 Dictionaries

Alcaraz, Enrique et al. (2006): Diccionario de términos del calzado e industrias afines: inglés-español, Spanish-English. Barcelona: Ariel.

Alcaraz, Enrique/Hughes, Brian/Campos, Miguel Ángel (1999): Diccionario de términos de marketing, publicidad y medios de comunicación: inglés-español, Spanish-English. Barcelona: Ariel.

Alcaraz, Enrique et al. (2000): Diccionario de términos de turismo y de ocio: inglés-español, Spanish-English. Barcelona: Ariel.

Alcaraz, Enrique/Hughes, Brian (2007): Diccionario de términos jurídicos/A dictionary of legal terms: inglés-español, Spanish-English. 10th ed. Barcelona: Ariel.

Alcaraz, Enrique/Martínez Linares, María Antonia (1997): Diccionario de Lingüística Moderna. Barcelona: Ariel.

Richards, Jack C./Platt, John/Platt, Heidi (1993): Longman dictionary of language teaching and applied linguistics. London: Longman.

9.2 Monographes and articles

Águila Escobar, Gonzalo. (2006). Las nuevas tecnologías al servicio de la lexicografía: Los diccionarios electrónicos. In: Villayandra, Milka (ed.): *Actas del XXXV Simposio Internacional de la Sociedad Española de Lingüística*. León: Universidad de León, 1–23. (also available at: [<http://www3.unileon.es/dp/dfh/SEL/actas.htm>, last access November 2, 2018]).

Alcaraz, Enrique/Mateo, José/Yus, Francisco (2007) (eds.): *Las lenguas profesionales y académicas*. Barcelona: Ariel.

Arntz, Reiner/Picht, Heribert (1989): *Introducción a la terminologia*. Madrid: Fundación Germán Sánches Ruipérez.

Asher, Ronald/Simpson, James M. Y. (eds.) (1994): *The encyclopaedia of language and linguistics*. Oxford: Pergamon Press.

Bisquerra, Rafael (1989): *Métodos de investigación educativa: Guía práctica*. Barcelona: CEAC.

Brown, James Dean (1988): *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.

Butler, Christopher (1985): *Statistics in Linguistics*. Oxford: Blackwell.

Cabré, M. Teresa (1999). *La terminología. Representación y comunicación*. Barcelona: Universitat Pompeu Fabra.

Celaya, María Luz (1992): *Transfer in English as a foreign language: A study on tenses*. Barcelona: PPU.

Cenoz, Jasone/Jessner, Ulrique (2000): *English in Europe: The acquisition of a third language*. Clevedon: Multilingual Matters.

Crystal, David (1980): *A dictionary of Linguistics and Phonetics*. London: Blackwell.

Crystal, David (1992): *An encyclopaedic dictionary of language and lan*guages. Oxford: Blackwell.

Crystal, David (1997): *English as a global language*. Cambridge: Cambridge University Press.

Crystal, David (2010): *The Cambridge encyclopaedia of language*. 3rd edition. Cambridge: Cambridge University Press.

Ellis, Rod (1994): *The study of second language acquisition*. Oxford: Oxford University Press.

Estaire, Sheila (1990): La programación de unidades didácticas a través de tareas. In: *Cable* 5, 28–39.

Howatt, Anthony P. R. (1984): *A history of English language teaching*. Oxford : Oxford University Press.

Johnson, Keith/Johnson, Helen (1998): *Encyclopedic dictionary of Applied Linguistics*. Oxford: Blackwell.

Krashen, Stephen D. (1981): *Second language acquisition and second language learning*. Oxford: Pergamon Press.

Lehmann, Christian (1996): Linguistische Terminologie als relationales Netz. In: Knobloch, Clemens/Schaeder, Burkhard (eds.): *Nomination - fachsprachlich und gemeinsprachlich*. Opladen: Westdeutscher Verlag, 215–267. (also available at: [<http://www.christianlehmann.eu/ling/ling_meth/ling_description/lexicography/terminol_dictionary.html>; last access: November 7, 2016].

Lewandowski, Theodor (1995): *Diccionario de Lingüística*. 4th edition. Madrid: Cátedra, D.L.

Martínez Linares, María Antonia (2007): Sobre la (morfo)sintaxis de las lenguas de especialidad. In: Alcaraz, Enrique/Mateo, José M./Yus Ramos, Francisco (eds.): *Las lenguas profesionales y académicas*. Barcelona: Ariel, 13–25.

McLaren, Neil/Madrid, Daniel (eds.) (1996): *A handbook for TEFL*. Alcoy: Marfil.

Morris, David W. H. (1988): *Dictionary of communication disorders*. London/Philadelphia: Whurr Publishers.

Muñoz, Carme (ed.) (2000): *Segundas lenguas: Adquisición en el aula*. Barcelona: Ariel.

Nicolosi, Lucille/Harryman, Elizabeth/Kresheck, Janet (1978): *Terminology of communication disorders*. Baltimore: Lippincott Williams & Wilkins.

Nortes, Andrés (1991): *Estadística teórica y aplicada*. Barcelona: DM/PPU.

Nunan, David (1988): *The learner-centred curriculum*. Cambridge: Cambridge University Press.

Nunan, David (1989): *Designing tasks for the communicative classroom*. Cambridge: Cambridge University Press.

Nunan, David (1992): *Research methods in language learning*. Cambridge: Cambridge University Press.

Peña Casanova, Jordi (1988): *Manual de logopedia*. Barcelona: Masson.

Renkema, Jan (1993): *Discourse studies: An introductory textbook*. Amsterdam: Benjamin's.

Ribé, Ramón/Vidal, Nuria (1995): *La enseñanza de la lengua extrajera en la Educación Secundaria: agenda práctica para aprender y enseñar una lengua extranjera en la ESO*. Madrid: Editorial Alhambra.

Salaberri, Sagrario (1999): *Lingüística aplicada a la enseñanza de lenguas extranjeras*. Almería: Universidad, Servicio de Publicaciones.

Spolsky, Bernard (1999): *Concise encyclopaedia of educational linguistics*. Amsterdam/New York : Elsevier.

Trask, Robert L. (1993): *A dictionary of grammatical terms in Linguistics*. London: Routledge.

Ur, Penny (1996): *A course in language teaching: Practice and theory*. Cambridge: Cambridge University Press.

Willis, Jane (1981): *Teaching English through English: A course in classroom language and techniques*. Harlow: Longman.

Woods, Anthony/Fletcher, Paul/Hughes, Arthur (1986): *Statistics in language studies*. Cambridge: Cambridge University Press.

Zanón, Javier (1993): *Claves para la enseñanza de la lengua extranjera: Primaria, 2º ciclo*. Madrid: MEC.

Zanón, Javier (1999): *La enseñanza del español mediante tareas*. Madrid: Edinumen.

Janusz Taborek

Multilingual LSP dictionary

Lexicographic conception of a dictionary of football language

Abstract: This paper deals with a lexicographical conception of a multilingual LSP dictionary as illustrated by a print dictionary of football language including four languages: Polish, Russian, English, and German –*Wörterbuch der Fußballsprache. Polnisch-Russisch-Englisch-Deutsch* (Taborek 2014), which contains more than 2,400 entries for each of the four languages. It intends to present a solution to typical questions in macrostructure, e.g. lemmatisation, multi-word-units and nominal forms, in microstructure, e.g. markers and usage-based examples, as well as mediostructure.

Keywords: multilingual lexicography, LSP lexicography, football language, sports linguistics

1 Introduction

The interest of lexicographers in the language of football as a Language for Special Purposes (LSP) results from the growing appeal of this sport in society (e.g. Zieliński 2002) as well as from the presence of football in the media. The popularity of football, a discipline developed among the working class in British cities, grew rapidly along with the medialisation and commercialization at the turn of the millennium. Also, the process of political, economic and social globalization as well as the common labour market in Europe in the 1980s and 1990s contributed to the internationalisation of football. Europe has still been the focal point of the football business, even though this sport has a long tradition in South America (e.g. Argentina, Brasil and Uruguay) and is gaining popularity in Africa, Asia and the USA. One of the consequences of globalisation and internationalization is multilingualism, because players and coaches in a football team speak different languages and English is not always the lingua franca in football. This is the first reason to treat the language of football as an object of linguistic inquiry. The second reason is that, at least since 1960s (e.g. George Best or later David Beckham, who were perceived as icons), football has been a part of pop culture. Lastly, football language and its phraseology are used in public language, e.g. *to kick something off* 'to begin or cause something to

Janusz Taborek: Adam Mickiewicz University in Poznań, al. Niepodległości 4, 61-874 Poznań, janusz.taborek@amu.edu.pl

https://doi.org/10.1515//9783110607659-012

begin', *to know the score* 'to be aware of the essential facts about a situation' or *to score an own goal* 'an act that unintentionally harms one's own interests' [<http://blog.oxforddictionaries.com>; last access: February 8, 2017].

2 The language of football – Linguistics – Lexicography

Linguists' interest in the language of football in the last decade is best exemplified by the number of volumes of papers on this topic, e.g. *The Linguistics of Football* (Lavric/Pisek/Skinner/Stadler 2008), *Flickflack, Foul und Tsukahara – der Sport und seine Sprache* (Burkhardt/Schlobinski 2009) or *Sprache und Fußball im Blickpunkt linguistischer Forschung* (Taborek/Tworek/Zieliński 2012) as well as studies on the language of football, e.g. Lewandowski (2013). The lexicographical registration of football lexicology has a longer tradition. Lipoński (2009: 32) mentions the English "Sportsman Dictionary" published in 1735 as the first dictionary of sports language. The oldest collection of football terminology in German seems to be the study by Konrad Koch, a German sports teacher, "Deutsche Kunstausdrücke des Fußballspiels" ('German expressions of the football game'), which was published in 1903. In that study, the first bilingual list of English-German terminology is presented (Koch 1903: 172). The dictionaries of sports, and especially football language, are usually published to herald sports events held in the same year, especially the Olympic Games, e.g. Wehlen (1972) for German, the European Championships, e.g. the English-German Elsevier-dictionary by Sirges (1980) or the World Cup, e.g. the multilingual English-German-French-Spanish dictionary (Binder/Brasse 1998). The last two decades have seen a growing number of dictionaries of football, especially in multilingual editions. The reasons for this trend are: the aforementioned popularity of football, medialization and digitalization as well as globalization and multilingual teams. Let us give some examples of multilingual dictionaries of football language published in this century that contain one of the four languages English, German, Polish and Russian: *PONS. Fachglossar Fußball* for French and German (Kerndter 2001), *Kauderwelsch Dictionary. Fußball* with more than 200 words in German, English, French, Italian, Spanish, Portuguese? and Turkish (Yıldırım 2006), *Англо-русский и русско-английский словарь футбольных терминов* for Russian and English (Zaharovich 2002), monolingual *Wörterbuch der Fußballsprache* (Burkhardt 2006), *Fußball. Wörterbuch deutsch-polnisch polnisch-deutsch* for German and Polish (Taborek 2006), *Lernwörterbuch der Fußballsprache* for German and French (Seelbach 2008), *Langenscheidt. UEFA. Praxiswörterbuch* for German, French and English (Simmen 2008 and the second edition 2010), *Das Wörterbuch der Fußballsprache* for Polish, Russian, English and German (Taborek 2014), English-Portuguese (Ramos/

Henry-Ramos 2014) and *Bendelow and Kidd's Dictionary of Football* (Bendelow/Kidd 2015).

3 Origins and structure of the dictionary

The dictionary of football language, *Das Wörterbuch der Fußballsprache Polnisch-Russisch-Englisch-Deutsch* (Taborek 2014), henceforth referred to as WFS, is a continuation of the author's earlier lexicographical work on football terminology and lexicology, of which the result was the bilingual German-Polish dictionary of football language (Taborek 2006). That dictionary was published on the occasion of the 2006 World Cup hosted by Germany. The bilingual dictionary included items excerpted from two major sources: dictionary entries in sports and football dictionaries and web corpus research. The project of the dictionary in question was started in 2008. The dictionary was designed from the beginning as a multilingual dictionary and includes four languages: Polish, German, English and Russian. The reason for the inclusion of the Russian language was the fact that Euro 2012 was co-hosted by Ukraine, where Ukrainian and Russian are spoken. Another rationale behind this decision was that the 2018 World Cup 2018 is scheduled to take place in Russia. Among the consultants of the WFS dictionary were native speakers of all four languages, linguists working on football language as well as amateur football players. As a result, the dictionary came to include more than 2,400 entries for each language, and is to my knowledge the largest multilingual football dictionary with regard to the number of lemmas in every language. The WFS dictionary was published in April 2014 on the occasion of the World Cup in Brazil.

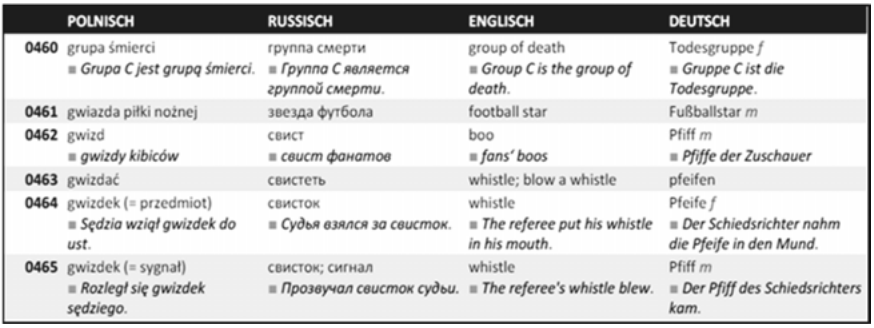
The structure of the dictionary is in accordance with the idea of Haensch (1991: 2923), cf. also Nielsen (1994), for multilingual print dictionaries of languages for special purposes.

> Das jeweilige Wörterbuch geht von einer Ausgangssprache aus, deren Wortschatz alphabetisch angeordnet ist, die Äquivalente in den übrigen Sprachen stehen daneben in mehreren Spalten. Der alphabetische Zugriff auf den Wortschatz der Ausgangssprache ist problemlos: dieser ist bereits alphabetisch geordnet, für die übrigen Sprachen gibt es hinter dem Hauptteil je ein alphabetisches Register mit Angabe einer Nummer, die jeweils vor dem Stichwort der Ausgangssprache steht, oder mit einer Seitenangabe. Dieses ist die häufigste und sicher auch die sinnvollste Anordnung eines alphabetischen Fachwörterbuches.
>
> Haensch 1991: 2923

According to Haensch, who calls this way of order "the most often and most useful order of an alphabetical LSP dictionary" (Haensch 1991: 2923), the starting point of the WFS dictionary is the Polish language and Polish terms are ordered alphabetically and have a forthcoming number. All equivalents and translations are given in columns, one column for every language. After the main part of the dictionary there is a register for English, German and Russian, where terms are ordered also alphabetically and include a number, which serves as a link to the entry in the main part of the dictionary.

The dictionary is a print dictionary, which results in specific outer texts, cf. Gouws (2004), Klosa (2009) and Bielińska (2010). The so-called frame structure of the WFS dictionary contains outer texts in Polish, English, German and Russian –a short multilingual preface with the objectives of this dictionary as well as the list of consultants and contributors and an introduction that includes a user manual as well as symbols and abbreviations used in the dictionary. The main part of the dictionary is the central list of 2,412 entries and has the following layout (WFS 2014: 47).

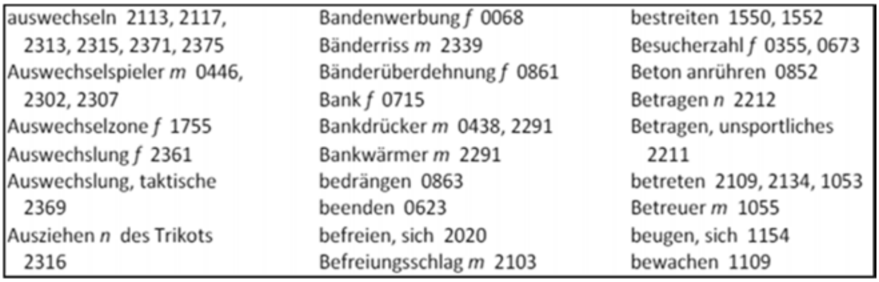
Tab. 1: The central list of entries in the dictionary (WFS 2014: 47)



	POLNISCH	RUSSISCH	ENGLISCH	DEUTSCH
0460	grupa śmierci ■ *Grupa C jest grupą śmierci.*	группа смерти ■ *Группа C является группой смерти.*	group of death ■ *Group C is the group of death.*	Todesgruppe *f* ■ *Gruppe C ist die Todesgruppe.*
0461	gwiazda piłki nożnej	звезда футбола	football star	Fußballstar *m*
0462	gwizd ■ *gwizdy kibiców*	свист ■ *свист фанатов*	boo ■ *fans' boos*	Pfiff *m* ■ *Pfiffe der Zuschauer*
0463	gwizdać	свистеть	whistle; blow a whistle	pfeifen
0464	gwizdek (= przedmiot) ■ *Sędzia wziął gwizdek do ust.*	свисток ■ *Судья взялся за свисток.*	whistle ■ *The referee put his whistle in his mouth.*	Pfeife *f* ■ *Der Schiedsrichter nahm die Pfeife in den Mund.*
0465	gwizdek (= sygnał) ■ *Rozległ się gwizdek sędziego.*	свисток; сигнал ■ *Прозвучал свисток судьи.*	whistle ■ *The referee's whistle blew.*	Pfiff *m* ■ *Der Pfiff des Schiedsrichters kam.*

The outer texts in the back matter are the three registers of terms in German, English and Russian. All terms are arranged in strict alphabetical order with references to the identification number of the entry in the central list of the dictionary and they are presented below (WFS 2014: 187).

Tab. 2: Register of German terms (WFS 2014: 187)



auswechseln 2113, 2117, 2313, 2315, 2371, 2375
Auswechselspieler *m* 0446, 2302, 2307
Auswechselzone *f* 1755
Auswechslung *f* 2361
Auswechslung, taktische 2369
Ausziehen *n* des Trikots 2316
Bandenwerbung *f* 0068
Bänderriss *m* 2339
Bänderüberdehnung *f* 0861
Bank *f* 0715
Bankdrücker *m* 0438, 2291
Bankwärmer *m* 2291
bedrängen 0863
beenden 0623
befreien, sich 2020
Befreiungsschlag *m* 2103
bestreiten 1550, 1552
Besucherzahl *f* 0355, 0673
Beton anrühren 0852
Betragen *n* 2212
Betragen, unsportliches 2211
betreten 2109, 2134, 1053
Betreuer *m* 1055
beugen, sich 1154
bewachen 1109

The front matter also includes a list of football language dictionaries which contain at least one of the four languages of the WFS dictionary. Those dictionaries were consulted during the lexicographic process and we make no claim to the completeness of the list[1].

4 Macrostructure of the dictionary

In this section, we aim to discuss and present the questions of (i) lemma selection, (ii) base form, i.e. the form of a lemma, (iii) multi-word units as lemmas.

The dictionary contains mainly special terms of football language. We took into account different genres, such as: official documents (e.g. match and competition regulations), print and online match reports, television, radio and minute-by-minute commentary, as well as the language of football fans used in Internet forums and fanzines. We even consulted dictionaries of fan language, e.g. Zaharovich (2002). Apart from that we could not fail to include terms from related sports disciplines or terms from other languages for special purposes, e.g. the language of sport (*match, win, lose*)[2], the language of medicine (*knee injury*), the language of law (*contract*) or media language (*live broadcast*).

The dictionary also lists some proper names, used as components of lemmas, when they are official names of competitions, e.g. *UEFA Cup*. The names of football clubs, national teams or player names are used only in lexicographic examples,

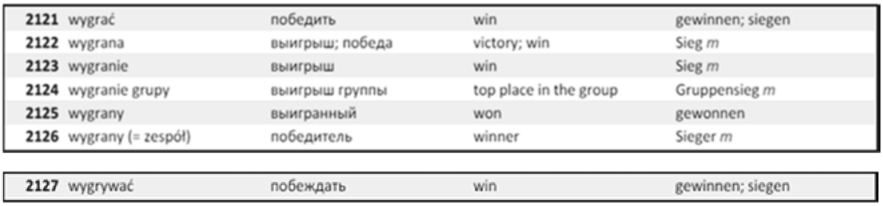
[1] For the bibliography of football linguistics including football dictionaries, cf. The Football and Language Bibliography Online of The Innsbruck Football Research Group at the University in Innsbruck, Austria [<https://www.uibk.ac.at/msp/projekte/sprache_fussball/bibliography/>; last access: February 8, 2017] as well as the list of the Sports Language Research Group (SpoLaReG) at Adam Mickiewicz University in Poznań [<http://spolareg.amu.edu.pl>; last access: February 8, 2017].

[2] The language of football is viewed as a part of the language of sport, which is seen as a part of a single language system, cf. Bergh/Ohlander (2012: 17).

which are designed to show the use or specific lexical or grammatical features of the lemma, e.g. *Inter is/are* (*at the*) *top of the table* in the entry *top of the table* (WFS 2014: 141), *France reached/entered the quarter final* in the entry *quarterfinal* (WFS 15) or *Advocaat cancelled his contract* in the entry *to cancel* (WFS 2014: 109). Dictionary entries also include eponyms: both official eponyms, e.g. *Bosman ruling*, and colloquial terms from football jargon, e.g. the Polish term *robinsonada* 'diving save'. Furthermore, there are also some terms that require encyclopaedic or expert knowledge, e.g. *hand of god* used for the goal scored illegally by Argentinian football player Diego Armando Maradona with his hand in a game against England during World Cup 1986 in Mexico.

Another issue linked with the base form of the lemma is the problem of languages with rich flection: two of the languages compared in the dictionary are Slavonic. The general rule for the form of the lemma is that the noun is given in the nominative and singular, with the exception of nouns having only plural forms, the so called *pluralia tantum*, or nouns used in plural e.g. *veterans*. The verb is given in the form of active infinitive and for the Polish and Russian languages, both perfective and imperfective forms are seen as lemmas. Hence, verbs having both forms, perfective and imperfective, occur twice in the dictionary, but it is on the one hand user friendly, and on the other hand this both forms do not necessarily follow each other in alphabetical order, see the example below for the Polish aspect pair *wygrać* – *wygrywać* 'to win', which are separated in alphabetical order –see the numbers (WFS 2014: 162.163).

Tab. 3: Perfective (number 2121) and imperfective (number 2127) verbs (WFS 2014: 162.163)



2121	wygrać	победить	win	gewinnen; siegen
2122	wygrana	выигрыш; победа	victory; win	Sieg *m*
2123	wygranie	выигрыш	win	Sieg *m*
2124	wygranie grupy	выигрыш группы	top place in the group	Gruppensieg *m*
2125	wygrany	выигранный	won	gewonnen
2126	wygrany (= zespół)	победитель	winner	Sieger *m*
2127	wygrywać	побеждать	win	gewinnen; siegen

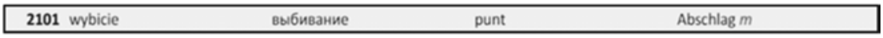
Adjectives, which are declinable in German, Russian and Polish, are presented in their undeclined form in German, e.g. *rotwürdig* 'sending-off offence'. In Russian and Polish adjectives are presented as lemmas with the default values of grammatical categories, i.e. nominative case, male gender, singular number and positive degree, e.g. *wyjazdowy* 'away' or *выиграный* 'won'. The remaining word classes, i.e. adverbs and preposition, occur as parts of multi-word units.

Multi-word units (MWU) are designed to be seen as lemmas in a LSP dictionary and this becomes clear when we compare terms in the languages of the WFS dictionary. The fact that the number of multi-word units varies across the languages is related to their morphology and spelling. Especially in German there is a possibility of combining words, or morphs, into complex units, which are spelled together, but also in German we have examples of multi-word units, e.g. *direkte Qualifikation* ‘direct qualification’. The number of MWUs in the Slavonic languages is greater and we can find e.g. in Polish MWUs with the nucleus (or head) represented by a noun, e.g. *atak pozycyjny* ‘positional attack’, by a verb, e.g. *założyć siatkę* ‘nutmeg, do a nutmeg’ or *wznawiać grę od środka boiska* ‘to kick off’, by an adjective or a participle, e.g. *niezgodny z przepisami* ‘illegal’, by a preposition, e.g. *u siebie* ‘home’, cf. the study of the Polish multi-word units in the WFS dictionary by Taborek (2015).

5 Microstructure in the dictionary

The microstructure in a multilingual LSP dictionary is designed to be limited to the minimum because of the specific lexical and grammatical features in the compared languages. Grammatical information is reduced in the central list of the WFS to the gender information for German nouns, *m* for masculine nouns, *f* for feminine nouns and *n* for neuter nouns, and they are presented as in the example below.

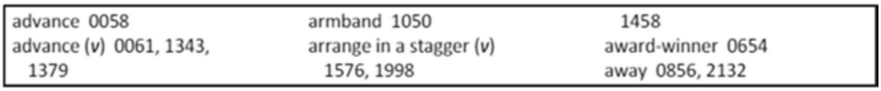
Tab. 4: Gender information for German nouns (WFS 2014: 160)



2101	wybicie	выбивание	punt	Abschlag *m*

Apart from the central list of the WFS, grammatical information is placed in the registers, where gender information for the German nouns is repeated, e.g. *Spielprotokoll n 1331*, and the part of speech is supplied in the English homographs to make distinctions between nouns and verbs, e.g. *advance* as a noun vs. *advance* as a verb, which is marked with the cursive *v* in brackets as seen below.

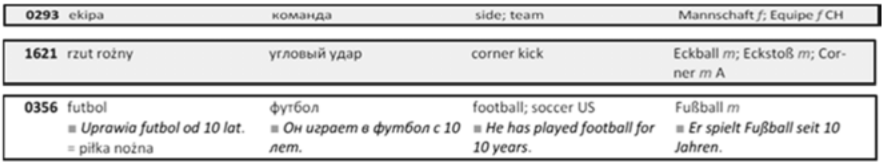
Tab. 5: Homograph nouns and verbs in the English register (WFS 2014: 205)



advance 0058	armband 1050	1458
advance (*v*) 0061, 1343,	arrange in a stagger (*v*)	award-winner 0654
1379	1576, 1998	away 0856, 2132

In the WFS we took into account the regional varieties in the language of football. Regional markers (*diatopische Markierung*) are used if the term is used only in Swiss German, which is marked with CH, in Austrian German, marked with A, and in American English, marked with US, as in the examples below.

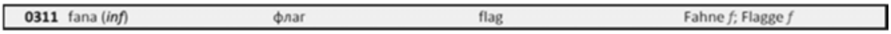
Tab. 6: Regional markers in the dictionary (WFS 2014: 35.127.40)



0293	ekipa	команда	side; team	Mannschaft *f*; Equipe *f* CH
1621	rzut rożny	угловый удар	corner kick	Eckball *m*; Eckstoß *m*; Corner *m* A
0356	futbol ■ ***Uprawia futbol od 10 lat.*** = piłka nożna	футбол ■ ***Он играет в футбол с 10 лет.***	football; soccer US ■ ***He has played football for 10 years.***	Fußball *m* ■ ***Er spielt Fußball seit 10 Jahren.***

Pragmatic markers (*pragmatische Markierung*) are used explicitly only to indicate informal usage, in which case the English standard abbreviation *inf* for informal is used, as in the example below in the entry for the Polish lemma *fana* 'flag', which is used only by football fans.

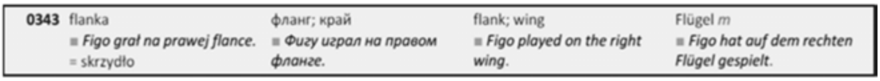
Tab. 7: Pragmatic markers in the dictionary (WFS 2014: 37)



0311	fana (*inf*)	флаг	flag	Fahne *f*; Flagge *f*

Semantic relations for the language of football were systematically well developed for English, German and French in the *Kicktionary* project[3], cf. Schmidt (2008), which includes semantic relations of synonyms, hypernyms, etc. as well as the concept of frames and scenes, in line with Charles Fillmore's approach. In contrast to *Kicktionary*, which is more of a lexical data base than a dictionary, we used only synonyms in the WFS. In the first language, in our case Polish, synonymy is indicated by an unambiguous sign of equivalency (=). In the other languages, the target languages in the central list of the WFS, the semicolon is used to separate synonyms. This seems to be the optimal solution in a print dictionary. The example below features synonyms *flanka* and *skrzydło* in the source language as well as synonyms *flank* and *wing* in one of the target languages.

Tab. 8: Synonyms in the dictionary (WFS 2014: 39)



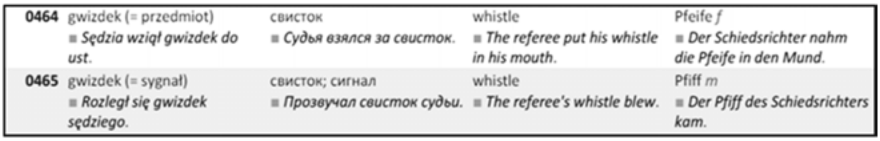
0343	flanka ■ ***Figo grał na prawej flance.*** = skrzydło	фланг; край ■ ***Фигу играл на правом фланге.***	flank; wing ■ ***Figo played on the right wing.***	Flügel *m* ■ ***Figo hat auf dem rechten Flügel gespielt.***

For the disambiguation of homonyms and polysemous words, e.g. *whistle* as 'a tubular wind instrument of wood, metal or other hard substance […]' (OED) and in the meaning 'an act of whistling' (OED), we use glosses, which contain usually the hypernym of the lemma as a genus proximum). The disambiguation concerns only

[3] [<http://www.kicktionary.de>; last access: February 8, 2017].

the source language, although in the example below (WFS 2014: 47), the equivalents in Russian and English are also ambiguous.

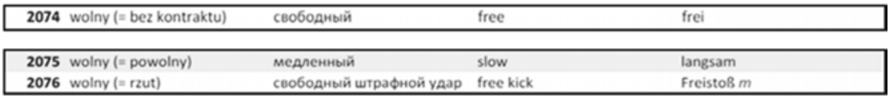
Tab. 9: Homonyms and polysemous words in the dictionary (WFS 2014: 47)



0464	gwizdek (= przedmiot) ■ *Sędzia wziął gwizdek do ust.*	свисток ■ *Судья взялся за свисток.*	whistle ■ *The referee put his whistle in his mouth.*	Pfeife *f* ■ *Der Schiedsrichter nahm die Pfeife in den Mund.*
0465	gwizdek (= sygnał) ■ *Rozległ się gwizdek sędziego.*	свисток; сигнал ■ *Прозвучал свисток судьи.*	whistle ■ *The referee's whistle blew.*	Pfiff *m* ■ *Der Pfiff des Schiedsrichters kam.*

The ambiguity in multilingual comparison can concern more difficult cases, even different parts of speech as in the example below presenting the Polish lemma *wolny*, which means 1. ‘free, without guilty player contract’, 2. ‘slow, not fast’ and 3. abbreviation for ‘free kick’. The ambiguity occurs only in one language while in other languages there is no need for disambiguation.

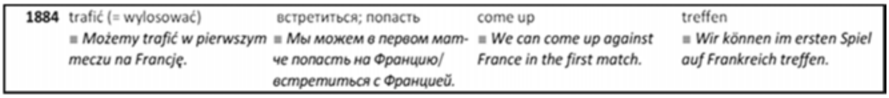
Tab. 10: Homographs in the dictionary (WFS 2014: 158.159)



2074	wolny (= bez kontraktu)	свободный	free	frei
2075	wolny (= powolny)	медленный	slow	langsam
2076	wolny (= rzut)	свободный штрафной удар	free kick	Freistoß *m*

The LSP dictionary in question contains not only equivalents for the three languages but also examples with its function to show the use of the lemma including its grammar and lexical properties. In the WFS we use examples for presenting syntagmatic relations of the lemma, its valency (colligation) and its collocations. Thanks to this approach, explicit grammatical features, e.g. the valency of the German verb *treffen* ‘to come up’, which collocates with the preposition *auf* and the accusative case (usually in dictionaries presented as *auf + Akk*), is shown implicitly in an example that shows grammatical properties, as exemplified below.

Tab. 11: Valency of the verbs in the dictionary (WFS 2014: 145)



1884	trafić (= wylosować) ■ *Możemy trafić w pierwszym meczu na Francję.*	встретиться; попасть ■ *Мы можем в первом матче попасть на Францию/ встретиться с Францией.*	come up ■ *We can come up against France in the first match.*	treffen ■ *Wir können im ersten Spiel auf Frankreich treffen.*

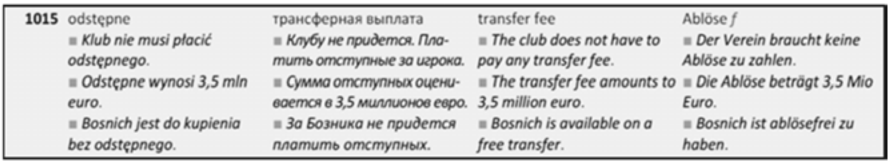
According to Seelbach (2008), we rely on well-formed sentences. The verb is used in the finite form instead of the infinitive and, if it is possible, the verb is used in the third person singular or plural present tense. This grammatical form is the most frequent verb form in live coverage in the traditional media like radio and television as well as in the modern media: on the Internet, in the most popular online sports genre, i.e. minute-by-minute reports. All examples used in the WFS dictionary are authentic sentences taken from the Internet and simplified, modified and generalized, e.g. by replacing proper names with generic terms like *team, player, opponent,* etc. For the reason of dictionary usability, we simplified two grammatical properties, (i) the aspect in Polish and Russian and (ii) the singular vs. plural dichotomy in English.

(i) In the Polish and Russian examples as well as lemmas (heads) only one aspect is used, either perfective or imperfective, depending on the context or frequency. In the example above (1884), the only possible aspect form in Polish and Russian is the perfective form trafić 'to come up' and not its imperfective form trafiać. The same goes for the Russian perfective verbs встретиться and попасть (their imperfective forms встречаться and попадать cannot be used in this context).

(ii) In line with standard British English usage, collective nouns denoting teams are followed by plural verbs only, e.g. Arsenal play... We do not use both verb forms which are in our opinion not necessary, e.g. Arsenal play/plays... or Arsenal play[s]...

(iii) The following dictionary entries show the solutions and the way of presenting lexical, grammatical and pragmatic features helping to use, understand or translate the given lemma.

Tab. 12: The use of the nouns in the dictionary (WFS 2014: 84)

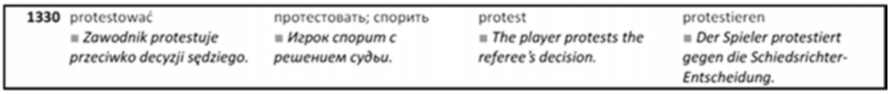


1015	odstępne	трансферная выплата	transfer fee	Ablöse *f*
	■ *Klub nie musi płacić odstępnego.*	■ *Клубу не придется. Платить отступные за игрока.*	■ *The club does not have to pay any transfer fee.*	■ *Der Verein braucht keine Ablöse zu zahlen.*
	■ *Odstępne wynosi 3,5 mln euro.*	■ *Сумма отступных оценивается в 3,5 миллионов евро.*	■ *The transfer fee amounts to 3,5 million euro.*	■ *Die Ablöse beträgt 3,5 Mio Euro.*
	■ *Bosnich jest do kupienia bez odstępnego.*	■ *За Бозника не придется платить отступных.*	■ *Bosnich is available on a free transfer.*	■ *Bosnich ist ablösefrei zu haben.*

The entry for the noun *transfer fee* (1015) contains three examples showing the use of the noun (i) as a grammatical object in a sentence, *The club does not have to pay any transfer fee*, (ii) as a grammatical subject in the example *The transfer fee amounts to 3,5 million euro* and (iii) as a part of an adverbial preposition phrase *without a transfer fee*, which is the translation of the Polish phrase *bez odstępnego*. In the English part we used the most common construction *to be available on a free transfer* (WFS 2014: 84).

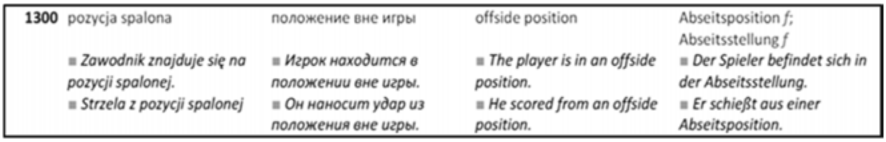
The complements (valency) of the verbs are also presented implicitly in examples as in entry 1330 for the verb *to protest*, which is followed by a direct object, but in the other compared languages there is a prepositional object with a language specific preposition: in Polish *protestować przeciwko* followed by the dative, Russian *спорить с* followed by the instrumental and German *protestieren gegen* followed by the accusative. The grammatical case that is required by the preposition or by the verb can be inferred from the example.

Tab. 13: The valency of the verbs in the dictionary (WFS 2014: 105)



1330	protestować ■ *Zawodnik protestuje przeciwko decyzji sędziego.*	протестовать; спорить ■ *Игрок спорит с решением судьи.*	protest ■ *The player protests the referee's decision.*	protestieren ■ *Der Spieler protestiert gegen die Schiedsrichter-Entscheidung.*

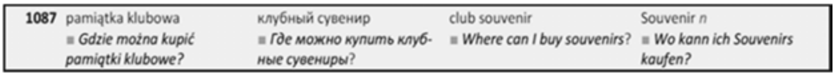
Tab. 14: Verbal collocations in the dictionary (WFS 2014: 103)



1300	pozycja spalona ■ *Zawodnik znajduje się na pozycji spalonej.* ■ *Strzela z pozycji spalonej*	положение вне игры ■ *Игрок находится в положении вне игры.* ■ *Он наносит удар из положения вне игры.*	offside position ■ *The player is in an offside position.* ■ *He scored from an offside position.*	Abseitsposition *f*; Abseitsstellung *f* ■ *Der Spieler befindet sich in der Abseitsstellung.* ■ *Er schießt aus einer Abseitsposition.*

Collocations are also presented in the examples usually in the form of complete sentences. Dictionary entry 1300 illustrates verbal collocations for the nominal collocation base *offside position* inclusive different preposition used in the prepositional phrase, i.e. *to be in an offside position* and *to score from an offside position* (WFS 2014: 103)[4].

Tab. 15: Textual, pragmatic and communicative properties in the dictionary (WFS 2014: 88)



1087	pamiątka klubowa ■ *Gdzie można kupić pamiątki klubowe?*	клубный сувенир ■ *Где можно купить клубные сувениры?*	club souvenir ■ *Where can I buy souvenirs?*	Souvenir *n* ■ *Wo kann ich Souvenirs kaufen?*

Regarding the lexico-grammatical features, some of the examples contain textual, pragmatic and communicative properties as in the dictionary entry below, where the noun phrase *club souvenir* is illustrated by the interrogative sentence *Where can I buy souvenirs*? (WFS 2014: 88), which can potentially be used during a football stadium visit.

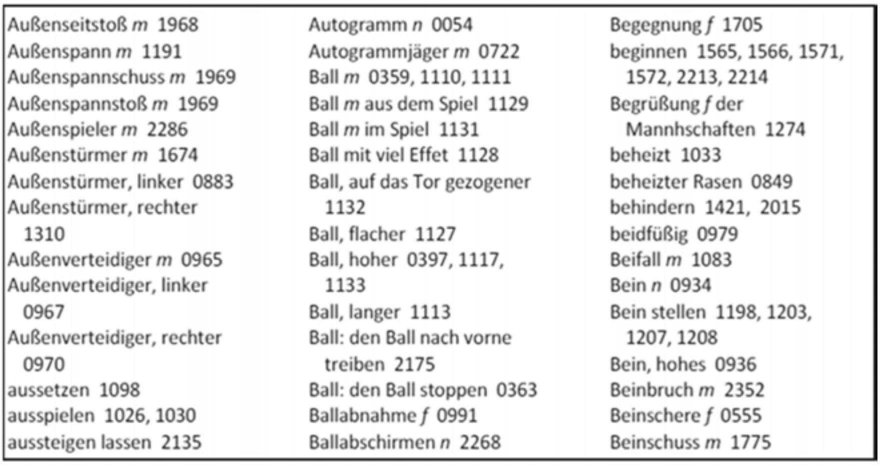
The aforementioned functions of the lexicographical examples, which illustrate the use of the lemma with regard to grammatical (e.g. valency), lexicological (e.g. collocations) and pragmatic (e.g. questions) properties are designed to be user-friendly. They should help users work efficiently with the dictionary as well as limit the metalanguage and the need for specialist linguistic knowledge.

[4] For the analysis of collocations in the WFS dictionary, cf. Kołupajło/Taborek (2016).

6 Mediostructure

The mediostructure in a print dictionary has to be reduced and there is no explicit reference system in the dictionary. In the central list, the synonyms have the function of referring to another dictionary entry, but the system concerns only source language, i.e. Polish. For the other languages, there is a reference system which links the lemma in the register via the number with the lemma in source language, as in the excerpt below, which shows the reference in German.

Tab. 16: The reference in German register (WSF 2014:187)



Außenseitstoß *m* 1968
Außenspann *m* 1191
Außenspannschuss *m* 1969
Außenspannstoß *m* 1969
Außenspieler *m* 2286
Außenstürmer *m* 1674
Außenstürmer, linker 0883
Außenstürmer, rechter 1310
Außenverteidiger *m* 0965
Außenverteidiger, linker 0967
Außenverteidiger, rechter 0970
aussetzen 1098
ausspielen 1026, 1030
aussteigen lassen 2135

Autogramm *n* 0054
Autogrammjäger *m* 0722
Ball *m* 0359, 1110, 1111
Ball *m* aus dem Spiel 1129
Ball *m* im Spiel 1131
Ball mit viel Effet 1128
Ball, auf das Tor gezogener 1132
Ball, flacher 1127
Ball, hoher 0397, 1117, 1133
Ball, langer 1113
Ball: den Ball nach vorne treiben 2175
Ball: den Ball stoppen 0363
Ballabnahme *f* 0991
Ballabschirmen *n* 2268

Begegnung *f* 1705
beginnen 1565, 1566, 1571, 1572, 2213, 2214
Begrüßung *f* der Mannhschaften 1274
beheizt 1033
beheizter Rasen 0849
behindern 1421, 2015
beidfüßig 0979
Beifall *m* 1083
Bein *n* 0934
Bein stellen 1198, 1203, 1207, 1208
Bein, hohes 0936
Beinbruch *m* 2352
Beinschere *f* 0555
Beinschuss *m* 1775

The excerpt from the dictionary register presented above shows also the way of presenting multi-word units in the register. In the print version of the dictionary multi-word units have to be listed twice or even more times. Multi-word units composed of adjectives and nouns are placed under the adjective, e.g. *langer Ball* ‘long pass’ as well as under *Ball, langer*, as in the example above. Verbal multi-word units are also presented under the noun, e.g. *Ball: den Ball stoppen*. This kind of multiplication can be omitted in an electronic version of the dictionary.

7 Conclusion

The multilingual LSP dictionary of football in question has been designed as a print dictionary, and thus it has to fulfil the requirements of the best possible multilingual dictionary, as postulated by Haensch (1991). Most of the football dictionaries published recently have an either thematic, e.g. Seelbach (2008), or a partially thematic and partially alphabetical order, e.g. *Langenscheidt UEFA Praxisw*örterbuch by Simmen (2008). An online football dictionary can combine both the thematic and the alphabetical order.

In conclusion, the following features of the WFS dictionary can prove useful to its user:

1. Many collocations are illustrated by examples of usage;
2. A great deal of grammatical information, e.g. complementation (valency) and irregular verb forms, is presented implicitly through examples;
3. Authentic examples of usage are presented as complete sentences, which can be used in communication, e.g. by learners or players;
4. Conjugated verb forms, usually given in the third person singular or plural in the present tense, are the most frequent verb forms in football live reports;
5. Examples of usage contain also textual, pragmatic and communicative features, e.g. questions;
6. Multi-word units are seen as lemmas and not as a combination of words or morphemes;
7. The dictionary includes three of the four most often taught languages in Europe (English, German, Russian) and can be used in foreign language instruction.

8 Bibliography

8.1 Dictionaries

Bendelow and Kidd's Dictionary of Football = Bendelow, Ian/Kidd, Jamie. Oakamoor: Bennion Kearny Limited, 2015.

Bilingual Dictionary of Football (*Soccer*) *Terms English/Portuguese and Portuguese/English – Dicionário Bilíngue de Termos de Futebol Inglês/Português e Português/Inglês* = Ramos, George Humberto/Henry-Ramos, Rhonda Abigail Bennett. Farasota: First Edition Design Publishing, 2014.

Elsevier's Football Dictionary: English-German/German-English = Sirges, Horst. London: Elsevier, 1980.

Football dictionary. Dictionnaire du football. Diccionario del fútbol. Fußball-Wörterbuch = Binder, Thomas/Brasse, Monika. Zürich: Fédération Internationale de Football Association, 1998.

Kauderwelsch Dictionary. Fußball – Football – Calcio – Fútbol – Futebol – Futbol in 7 Sprachen. Deutsch – English – Français – Italiano. Español – Português – Türkçe = Yildirim, Kaya. Bielefeld: Reise-Know-How Verlag Peter Rump, 2006.

Langenscheidt UEFA. Praxiswörterbuch Fußball. Deutsch – Englisch – Französisch = Simmen, Florian. Berlin, München, Wien, Zürich, New York: Langenscheidt/UEFA, 2008.

Lernwörterbuch der Fußballsprache. Deutsch-Französisch Französisch-Deutsch = Seelbach, Dieter. Hamburg: Buske, 2008.

OED = Oxford English Dictionary [<http://www.oed.com>; last access: February 8, 2017].

Piłka nożna. Słownik niemiecko-polski polsko-niemiecki. Fußball. Das Wörterbuch Deutsch-Polnisch Polnisch-Deutsch = Taborek, Janusz. Zielona Góra: Wydawnictwo Kanion, 2006.

PONS. Fachglossar Fußball. Französisch-Deutsch Deutsch-Französisch = Kerndter, Fritz. Stuttgart: Ernst Klett Verlag, 2001.

Regeln und Sprache des Sports. 2 Teile = Wehlen, Rainer. Mannheim, Wien, Zürich: Bibliographisches Institut/Dudenverlag, 1972.

WFS = *Das Wörterbuch der Fußballsprache. Polnisch – Russisch – Englisch – Deutsch* = Taborek, Janusz. Hamburg: Verlag Dr. Kovač, 2014.

Wörterbuch der Fußballsprache = Burkhardt, Armin. Braunschweig: Die Werkstatt Verlag, 2006.

Zaharovich = *Англо-русский и русско-английский словарь футбольных терминов*. Захарович, Л. А. Москва: Издательство АСТ, 2002.

8.2 Monographies and articles

Bergh, Gunnar/Ohlander, Sölve (2012): *Free kicks, dribblers* and *WAGs*. Exploring the language of “the peoples game”. In: *Moderna språk* 1, 11–46.

Bielińska, Monika (2010): *Lexikographische Metatexte. Eine Untersuchung nichintegrierter Außentexte in einsprachigen Wörterbüchern des Deutschen als Fremdsprache*. Frankfurt am Main et al.: Peter Lang.

Burkhard, Amin/Schlobinski, Peter (eds.) (2009): *Duden Thema Deutsch 10. Flickflack, Foul und Tsukahara. Der Sport und seine Sprache.* Mannheim: Bibliographisches Institut.

Gouws, Rufus H. (2004): Outer texts in bilingual dictionaries. In: *Lexikos* 14, 67–88.

Haensch Günther (1991): Die mehrsprachigen Wörterbücher und ihre Probleme. In: Hausmann, Franz Joseph/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.): *Wörterbücher. Dictionaries. Dictionnaires*. HSK 5.3., Berlin/New York: De Gruyter, 2909–2927.

Klosa, Anette (2009): Außentexte in elektronischen Wörterbuchern. In: Beijk, Ebert et al. (eds.): *Fons Verborum Feestbundel Fons Moerdijk*. Amsterdam: Gopher BV, 49–60.

Koch, Konrad (1903): Deutsche Kunstausdrücke des Fußballspieles. In: Streicher, Oskar (ed.): *Zeitschrift des allgemeinen deutschen Sprachvereins*. XVIII (6). Berlin: Verlag des Allgemeinen Sprachvereins (F. Berggold), 169–172.

Kołupajło, Kornelia/Taborek, Janusz (2016): Kollokationen in einem mehrsprachigen Fachwörterbuch. Dargestellt am Beispiel des Wörterbuches der Fußballsprache. In: Bąk, Paweł/Rolek, Bogusława (eds.): *Vom Wort zum Gebrauch. Wortbedeutung und ihre Eingebundenheit in Diskurse.* Frankfurt am Main et al.: Peter Lang, 157–169.

Lavric, Eva/Pisek, Gerhard/Skinner, Andrew/Stadler, Wolfgang (eds.) (2008): *The Linguistics of Football*. Tübingen: Gunter Narr Verlag.

Lewandowski, Marcin (2013): *The Language of Football: an English-Polish Contrastive Study*. Poznań: Wydawnictwo Naukowe UAM.

Lipoński, Wojciech (2009): “Hey, ref! Go, milk the canaries!” On the distinctiveness of the language of sport. In: *Studies in Physical Culture and Tourism*. 16 (1). Special issue. Sports language & linguistics, 19–36.

Nielsen, Sandro (1994): *The Bilingual LSP Dictionary: Principles and Practice for Legal Language*, Tübingen: Narr.

Schmidt, Thomas (2008): The Kicktionary: Combining corpus linguistics and lexical semantics for a multilingual football dictionary. In: Lavric, Eva/Pisek, Gerhard/Skinner, Andrew/Stadler, Wolfgang (eds.): *The linguistics of football*. Tübingen: Gunter Narr Verlag, 11–22.

Taborek, Janusz (2015): Die Struktur polnischer Mehrworteinheiten in einem mehrsprachigen Fachwörterbuch. In: *Linguistische Treffen in Wrocław* 11, 243–250.

Taborek, Janusz/Tworek, Artur/Zieliński, Lech (eds.) (2012): *Sprache und Fußball im Blickpunkt linguistischer Forschung*. Hamburg: Verlag Dr. Kovač.

Zieliński, Lech (2002): Wpływ słownictwa sportowego na język polityki. In: Szpila, Grzegorz (ed.): *Język a komunikacja 4, Bd. I Język trzeciego tysiąclecia II Nowe oblicza komunikacji we współczesnej polszczyźnie*. Kraków: Tertium, 259–272.