

# Science, Systemic Functional Linguistics and Language Change

*A Festschrift for David Banks*



Edited by

Shirley Carter-Thomas  
and Clive E. Hamilton

# Science, Systemic Functional Linguistics and Language Change



# Science, Systemic Functional Linguistics and Language Change:

*A Festschrift for David Banks*

Edited by

Shirley Carter-Thomas  
and Clive E. Hamilton

**Cambridge  
Scholars  
Publishing**



Science, Systemic Functional Linguistics and Language Change:  
A Festschrift for David Banks

Edited by Shirley Carter-Thomas, Clive E. Hamilton

This book first published 2019

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

Copyright © 2019 by Shirley Carter-Thomas, Clive E. Hamilton and contributors

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-5275-3908-3

ISBN (13): 978-1-5275-3908-2

# TABLE OF CONTENTS

List of Contributors .....	vii
Editors' introduction.....	1
Shirley Carter-Thomas and Clive Hamilton	
<b>Part I: Languages for specific purposes</b>	
Chapter One.....	11
The language of science and the native speaker: Nature, nurture, and norms	
Carmela Chateau Smith	
Chapter Two .....	41
Lexical density in scientific research genres	
Elizabeth Rowley-Jolivet	
Chapter Three .....	67
Imprégnation de la langue par la culture : l'exemple de l'anglais médical	
Pascaline Faure	
<b>Part II: Diachrony and science</b>	
Chapter Four.....	87
Diachronic research: A methodology for revealing a discipline's implicit frame of indexicality	
Dacia Dressen-Hammouda	
Chapter Five .....	107
The popularization of science in 19th century France: The "Bibliothèque utile" or the useful library	
Janet Ormrod	

Chapter Six.....	121
Metadiscourse and stance-taking in prefaces: A diachronic analysis Mohamed Saki	
Chapter Seven.....	139
Les définitions du mot « cancer » dans les dictionnaires et les traités médicaux : Approche diachronique Ghislaine Rolland-Lozachmeur	
<b>Part III: SFL approaches to language and language change</b>	
Chapter Eight.....	163
“We shall shortly be arriving into Reading”: A case study in contemporary language change Gordon Tucker	
Chapter Nine.....	183
When leaving is a thing: Construing Britain’s exit from the European Union Lise Fontaine	
Chapter Ten.....	201
Langue naturelle, langage iconique, langage symbolique Anca Christine Pascu	
Chapter Eleven.....	227
L’évolution pédagogique de la langue des linguistes allemands au cours du 19 <sup>e</sup> siècle Jacques François	

## LIST OF CONTRIBUTORS

**Shirley Carter-Thomas:** Institut Mines-Télécom Business School

**Carmela Chateau Smith:** Université de Bourgogne

**Dacia Dressen-Hammouda:** Université Clermont Auvergne

**Pascaline Faure:** Sorbonne Université

**Lise Fontaine:** University of Cardiff

**Jacques François:** Université de Caen Normandie

**Clive Hamilton:** Université de Paris

**Janet Ormrod:** IMT Atlantique (ex Télécom Bretagne)

**Anca Christine Pascu:** Université de Bretagne Occidentale

**Ghislaine Rolland-Lozachmeur:** Université de Bretagne Occidentale

**Elizabeth Rowley-Jolivet:** Université d'Orléans

**Mohamed Saki:** Université de Bretagne Occidentale

**Gordon Tucker:** University of Cardiff





# INTRODUCTION

SHIRLEY CARTER-THOMAS  
AND CLIVE HAMILTON

## 1. David Banks: Introductory remarks

David Banks has been a leading figure in linguistics, both within France and internationally, for a number of years. He is the author of a prolific body of research which has influenced several generations of researchers, and has also been successful in bringing together linguists of different theoretical backgrounds, languages and research domains. Emeritus Professor of English linguistics at *Université de Bretagne Occidentale* in Brest, where he has been based for over three decades, David Banks is originally from Newcastle-upon-Tyne. He has however been living outside the UK since 1975, initially in Iraq, and subsequently in France. He holds a degree in philosophy from the University of Cambridge, a doctorate in English linguistics from the *Université de Nantes*, and an HDR (*Habilitation à Diriger des Recherches*) from the *Université de Bordeaux 2*.

David Banks founded the French systemic functional linguistics group AFLSF (*Association Française de la Linguistique Systémique Fonctionnelle*), and was its Chairman for fourteen years. He was also formerly Head of the English department and Director of the ERLA (*Equipe de Recherche en Linguistique Appliquée*), the applied linguistics team at *Université de Bretagne Occidentale*, and organised regular conferences at his university, most of which gave rise to collective publications on a vast array of topics. The different papers assembled in this book by fellow scholars and colleagues of David Banks are intended as a tribute to these and other achievements, in recognition of the seminal role David Banks has played in linguistics. It is also hoped that other younger colleagues and scholars less familiar with his work will take an interest in the issues raised and be encouraged to delve further.

The title of the present book, “Science, SFL and Language Change: A Festschrift in honour of David Banks”, refers to three of the main themes that have run through his work over the years: scientific language, Systemic

Functional Linguistics (SFL), and diachrony and language change. Although these three strands have often been amalgamated in David Banks's work, we propose to briefly describe his interest in the three separately, before going on to outline the organisation of the book.

## 2. Science, SFL and Language Change

One of David Banks's main passions has been for the language of science, culminating in the publication of *The Development of Scientific Writing: Linguistic Features and Historical Context* (Banks 2008), for which he received the European Society for the Study of English Language and Linguistics Book Award in 2010. This work, after first discussing previous studies of the development of scientific language, including Halliday's exploration of the topic (Halliday 1988), traces the development of scientific writing as a genre, in terms of its linguistic features, from Chaucer's "Treatise on the Astrolabe" to a corpus of texts taken from the *Philosophical Transactions of the Royal Society*, covering the period 1700 to 1980. Basing his analysis on a close study of a range of linguistic features (first person pronouns, nominalisation, grammatical metaphor, passive forms and thematic structure), his conclusions point to some different patterns of development in the physical and biological sciences, such as the increased influence of mathematical modelling on texts in the physical sciences from the late 19th century onwards.

David Banks's focus on the grammatical features of scientific writing is a prominent trait of much of his work. In his earlier work, this emphasis was also tightly bound up with the pedagogical concern of providing a detailed description of scientific journal articles for the training of non-Anglophone scientists. Using a small set of articles in oceanography (1991, 1993, 1995 *inter alia*), David Banks surveyed a variety of grammatical features such as the use of personal pronouns, agentless passives, hedging patterns, etc. Rather than considering scientific texts from a rhetorical and sociolinguistic viewpoint, his starting point here was the text's linguistic specificities, those present as well as those absent (cf. Banks 1995).

David Banks's interest in applied linguistics in France can also be seen in his consistent support for the GERAS<sup>1</sup> conferences over the years, his membership of the editorial board of the *ASP* journal and his editorship (in the pre-Internet days) of the ESP France newsletter. He has done a great deal to promote LSP and ESP studies in the French linguistics landscape,

---

<sup>1</sup> *Groupe d'Etude et de Recherche en Anglais de Spécialité* (English for Specific Purposes Study and Research Group).

not only through his own work but also through his PhD supervision, participation in committees and the support he has always so generously provided to his colleagues.

The promotion of Systemic Functional Linguistics in France is also among one of David Banks's most significant achievements. As well as teaching SFL in his courses, disseminating SFL through his journal articles (cf. for example Banks 2002) and his recent book (2019), he is the author of an abridged SFL grammar of English published in French (2005), *Introduction à la linguistique systémique fonctionnelle de l'anglais*. With a foreword by Michael Halliday, this very readable grammar covers several aspects of SFL theory such as transitivity analyses, thematic structure and the interpersonal metafunction, and has been instrumental in enabling French linguists (students and teachers alike) to gain a better understanding of SFL theory. David Banks has in addition co-edited books bringing together SFL analyses of the French language on various features (2009, 2016). In an effort to build bridges between theories, he has also drawn parallels between SFL and the French theory of enunciative operations, based on the work of Antoine Culioli (Banks 2004).

The study of language as it develops over time is perhaps the linguistic approach that has come to best represent the work of David Banks, particularly when combined with the study of scientific discourse. In recent years David Banks has pioneered a diachronic approach to the study of English (and other languages) for specific purposes, or "diachronic ESP" (Banks 2005b; 2012; 2016). He argues that "understanding the way specialised texts have developed can help us understand why they are the way they are today" (Banks 2012, 55). Scientific language necessarily changes in function of the context of scientific research in which it is produced, in order to meet the communicative needs of its practitioners.

In another recent book on the birth of the scientific article (Banks 2017), he uses the method to compare and contrast the development of research articles taken from the first two academic periodicals in Britain and France, from their creation in 1665 until the end of the seventeenth century. The focus on the historical context of the late seventeenth century in the two countries enables him to place the texts solidly in the context from which they derived. It is firstly shown how this context influenced the different editorial decisions made by the respective editors of the *Journal des Sçavans* and the *Philosophical Transactions*, which led in turn to the texts they published also displaying differing linguistic features. David Banks makes a convincing case for the contrastive analysis of various linguistic features studied in the two journals, such as type of theme, nominalisations,

expressions of modality or process type, being directly explicable in terms of the historical context (Banks 2012).

The emphasis on three main themes in David Banks work should not be taken as an indication that his interests have only been limited to these topics. In addition to scientific texts, he has also analysed the language of the Bible, fiction, and the press (cf. Banks 1999). His work embraces lexicogrammar and the study of text, both from synchronic and diachronic viewpoints. He has also edited volumes on a variety of subjects: coordination and subordination, tense and aspect, poetic language, political texts and propaganda (cf. Banks 2005c; 2008).

### 3. Presentation of Contributions

The contributions in this book were chosen to provide a picture of the main fields of interest of David Banks, as outlined in section 2: SFL, the language of science and language change. The perspective offered is thus quite unique in the coverage it proposes of these three areas, which are rarely treated together. The eleven chapters can all be read separately, but the numerous points of convergence and inter-connections between the chapters and sections contribute to the book's overall coherence.

The authors are experienced linguists and/or LSP practitioners, all with considerable experience of teaching and research in higher education in France and the UK. Some are colleagues of David Banks from *Université de Bretagne Occidentale* in Brest, others are colleagues from the various scholarly associations David Banks have been associated with, or who have met him regularly in research contexts over the years. Several of the contributors have worked in the French university system and the book could therefore be of particular interest for scholars working in France or other European contexts, or for those with interests in contrastive linguistics. Four of the eleven chapters are in French.

The papers proposed in the first section all consider aspects of ESP in scientific discourse. **Carmela Chateau Smith** (Université de Bourgogne) focuses on the predominance of English as the language of scientific communication. After exploring the history of English in this role, she discusses some of the problematic issues linked to the notions of “native” and “non-nativeness” in academic writing in the sciences. The article ends with some suggestions for how language assistance can be best optimised by ESP teachers. Taking also scientific discourse as her starting point, **Elizabeth Rowley-Jolivet** (Université d'Orléans) focuses on the question of lexical density. She shows that although lexical density and the associated informational load are generally considered to be very high in scientific



discourse (cf. Halliday 1985), it is also important to take into account other factors, such as genre and mode. The analysis of a corpus comparing seven scientific research genres - two spoken (three-minute thesis presentations, conference talks), and five written (research articles, thesis and article abstracts, scholarly journal editorials, laboratory protocols, conference slides) - supplemented by analysis of their verbal density and subject pronoun use, produces results that show great variation. Whereas for the two speech genres lexical density stands at around 48%, it rises to 75% in writing, and also displays frequent correlations with the other two features examined. **Pascaline Faure** (Sorbonne Université) focuses in her chapter on medical English. She shows how English medical terminology is motivated on the one hand by the nature and constraints of Western medical practice (brevity and medical secrecy) and on the other hand by Anglo-Saxon culture (merit, science, travel, humour, food, and lifestyle). After providing examples to illustrate these two main influences, she goes on to consider the case of recently created medical terminology linked to the emerging field of 4P (or preventive) medicine, and to the possible influence of new types of medical cultures associated with this movement.

The second section of the book focuses on diachronic approaches in the analysis of specialised discourse. In her chapter, **Dacia Dressen-Hammouda** (Université Clermont Auvergne) explains the importance of diachronic analysis as a methodology for identifying and exploring the implicit content of disciplinary writing. Drawing on previous work carried out in the discipline of geology (Dressen-Hammouda, 2014), she shows that by situating current text features within past contexts, it is possible to identify implicit or occluded features of contexts that have become less explicit over time. The author claims that this socio-historical insight into the implicit features of writing has a considerable contribution to make in the field of teaching methods in specialised writing. **Janet Ormrod** (ex-Telecom Bretagne) focuses on the popularisation of science in 19<sup>th</sup> century France. Basing her analysis on a text intended for a lay audience (from the “Bibliothèque utile”) with an article on same topic that appeared in the scientific journal, *Journal des Sçavans*, her study identifies some of the characteristics that distinguish the two genres: “low science” and “high science. **Mohamed Saki** (Université de Bretagne Occidentale) in his chapter examines the interpersonal and evaluative dimensions of metadiscourse in six auctorial prefaces to English novels written in the 18<sup>th</sup>, 19<sup>th</sup> and 20<sup>th</sup> centuries. The analysis seeks to shed light on the permanent and changing characteristics of the uses of metadiscourse in the preface subgenre. **Ghislaine Rolland-Lozachmeur** (Université de Bretagne Occidentale) turns her attention to the medical domain (cf. chapter by Pascaline Faure

above). Her contribution analyses the definitions of the word “cancer” in language dictionaries and medical works, from the 18th century up to today’s health advice and information websites. The analysis suggests a great deal of consistency in the way the term is used, with the disease being frequently associated with a lexicon of combat.

The chapters in the third section, as well as focusing on language change, also raise issues related to linguistic theory. SFL approaches to language and language change form the basis of the chapters by **Lise Fontaine** (University of Cardiff) and **Gordon Tucker** (University of Cardiff). Starting from the premise that seeds of potentially permanent language change can be observed over the course of just a few years, Gordon Tucker focuses on a case of possible contemporary language change. The paper charts evidence, in corpus data, for the recent increase in the use of the preposition “into”, instead of “in”, as in the expression “We shall shortly be arriving into Reading” and offers possible explanations of why this apparent change is taking place. Lise Fontaine in her contribution analyses the linguistic complexity of the very topical neologism, *Brexit*. As a term, *Brexit* represents a complex nominal since it exploits nominal grammatical resources whilst expressing process semantics. She argues that while the newness of *Brexit* prevents the concept of (in)congruence from being operational, the SFL concept of construal in the other hand offers a potentially useful approach for empirical evidence of its semantics, with its construal being that of a hypothetical event with negative semantic prosody. **Anca Christine Pascu** (Université de Bretagne Occidentale) analyses the relationship between natural, iconic and symbolic languages. After tracing the development of linguistics and logic as disciplines, she shows how the distinction between them can also be used as a parallel to distinguish iconic language and symbolic language. Drawing on her background in computer science and natural language processing, her aim is to establish some characteristics of these three types of language and their combined impact on the construction of knowledge. In the final chapter, **Jacques François**, (Université de Caen Normandie) studies the evolution of 19th century German linguistic discourse, based on a close analysis of three linguists. In the first half of the century, illustrated by the writings of Wilhelm von Humboldt and Franz Bopp, the linguistic discourse is directed at peers. However, by the end of the century, with the rapid expansion of universities in Germany following the country’s unification in 1870, this discourse takes on a more applied educational dimension, illustrated in the excerpt from Meyer-Lübke. By distinguishing between different types of relationships between clauses, the author shows how SFL is a useful tool for comparatively assessing how the syntactic structure of these discourses evolved.

\*

\*\*

It remains for us to hope that readers will enjoy reading the contributions, which reflect the great respect in which David Banks is held, as much as we have enjoyed preparing this volume as a tribute to his achievements.

## References

- Banks, David. 1991. "Some observations concerning transitivity and modality in scientific writing." *Language Sciences* 13(1): 59–78.
- Banks, David. 1993. *Writing in Water. Aspects of the Scientific Journal Article*. Brest: E.R.L.A. Université de Bretagne Occidentale.
- Banks, David. 1995. "There is a cleft in your sentence: Less common clause structures in scientific writing." *Asp. 7-10*: 3-11.  
<https://journals.openedition.org/asp/370>
- Banks, David. 1999. "Decoding the information structure of journalistic clefts" *Interface: Journal of Applied Linguistics*, 14.1: 3-24.
- Banks, David. 2002. "Systemic Functional Linguistics as a model for text analysis." *Asp*, 35-36: 23-34. <https://journals.openedition.org/asp/1584>
- Banks, David. 2004. "Anglophone systemicists and French enunciativists: Shall the twain never meet?" *Language Sciences* 26(4): 391-410
- Banks, David. 2005a, *Introduction à la linguistique systémique fonctionnelle de l'anglais*. Paris: L'Harmattan.
- Banks, David. 2005b. "On the historical origins of nominalized process in scientific text." *English for Specific Purposes* 24/3, 347–367.
- Banks, David (Coord.) 2005c. *Aspects linguistiques du texte de propagande*, Paris, L'Harmattan.
- Banks, David. 2008. *The Development of Scientific Writing: Linguistic Features and Historical Context*. London: Equinox.
- Banks, David (Coord.) 2008. *La Langue, la linguistique et le texte religieux*. Paris, L'Harmattan.
- Banks, David, Simon Eason, Janet Ormrod (Eds.) 2009. *La linguistique systémique fonctionnelle et la langue française*. Paris, L'Harmattan.
- Banks, David. 2012. "Diachronic ESP: at the interface of linguistics and cultural studies", *Asp*, 61: 55-70.
- Banks, David. 2016. "Diachronic aspects of ESP." *Asp*, 69: 97-112.  
<https://journals.openedition.org/asp/4812>
- Banks, David, Ormrod, Janet (Coord.) 2016. *Nouvelles études sur la transitivité en français*. France. L'Harmattan.



- Banks, David. 2017. *The Birth of the Academic Article: Le Journal des Sçavans and the Philosophical Transactions, 1665-1700*. Sheffield: Equinox
- Banks, David. 2019. *A Systemic Functional Grammar of English. A simple introduction*. New York & London, Routledge.
- Dressen-Hammouda, Dacia. 2014. "Measuring the Voice of Disciplinarity in Scientific Writing: A Longitudinal Exploration of Experienced Writers in Geology." *English for Specific Purposes* 34: 14–25.
- Halliday, Michael A.K. 1985. *Spoken and written language*. Oxford: Oxford University Press.
- Halliday, M.A.K. 1988. "On the language of physical science". In M. Ghadessy. (Ed.), *Registers of Written English: Situational factors and linguistic features*. Pinter: London, 162–178. [reprinted in Halliday, M.A.K. & J.R. Martin 1993. *Writing Science. Literacy and discursive power*. The Falmer Press: London, 54–68].

**PART I:**  
**LANGUAGES FOR SPECIFIC PURPOSES**



# CHAPTER ONE

## THE LANGUAGE OF SCIENCE AND THE NATIVE SPEAKER: NATURE, NURTURE, AND NORMS

CARMELA CHATEAU SMITH

### Introduction

Two frequent themes in the work of David Banks are the importance of diachronic studies in English for Specific Purposes (ESP), and the difference in worldview between English and French. This study combines these themes by exploring the history of English as an international language of scientific communication, together with the notion of “native speaker”.

In the early twenty-first century, English is acknowledged as the international language of science, yet this predominance is comparatively recent. The first section of this study, the Nature of scientific language, traces the history of languages used for international communication. The second section, Nurture, explores a pilot corpus of 27 articles providing advice on effective writing for publication. The final section, Norms, examines aspects of linguistic monopoly in academia today, through a case study based on reviewers’ remarks. ●Optimising linguistic assistance for all, from first-year students to experienced reviewers, remains a key goal for ESP teachers within the scientific community.

### 1. Nature: the language of science over time

The history of English as an international language for scientific communication is linked to the history of science. Isaac Newton (1643–1727) was a “natural philosopher”, as underlined by David Banks in his comparison of Newton’s *Opticks* and Christiaan Huygens’ *Traité de la Lumière*, an exploration of “how scientific English came into being in the

late seventeenth century” (Banks, 2004: 2). The term “scientist” was coined much later, in 1834, by William Whewell (Ross, 1962: 71).

This section is divided into four parts, covering the impact of Greek and Latin until the late seventeenth century, then French in eighteenth- and nineteenth-century Europe, followed by Esperanto and the quest for a neutral international language in the early twentieth century, and finally the place of English in twenty-first century academia.

## 1.1 The influence of Greek and Latin

Many languages have spread beyond national borders, but Greek is “the European language with the longest recorded history” (Horrocks, 2010: 10). Much modern scientific vocabulary is based on Greek roots, by direct imports, and by imports from Greek via Latin, so that “over 60 percent of all English words have Greek or Latin roots; in the vocabulary of the sciences and technology, the figure rises to over 90 percent.” (Green, 2015: 12). Similar meanings are expressed by words of different origin, e.g. Greek *gé* and *logos* are found in “geological”; Latin *terra* produces “terrestrial”; Old English *eorþe* gives “earthly”. For other words (e.g. *agony*), “The route of transmission is from Greek to Latin to French.” (Stockwell & Minkova, 2001: 51).

In Britain, the use of Latin began during Roman occupation (AD 43 to AD 410), and continued long after the fall of the Roman Empire, notably as the language of the Catholic Church. Elsewhere in Europe, Latin was chosen in 1479 as the language for the newly founded University of Copenhagen, since its purpose was “essentially religious” (Mortensen & Haberland, 2012: 177). Although John Wycliffe’s English Bible was hand-printed as early as 1382, the Church of England’s fundamental English texts first appeared during the Tudor Reformation (*Book of Common Prayer*, 1569; *King James Bible*, 1611)

The Scientific Revolution, during the Stuart period (1603–1714), saw the discoveries of Kepler, Galileo, Harvey, Hooke, Boyle, and Newton, together with the works of Bacon, Descartes, Hobbes, and Locke, amongst others. In Lyell’s *Principles of Geology*, a footnote about Lazarro Moro’s *Sui Crostacei ed altri Corpi marini che si trovano sui Monti* (1740) underlines the difficulty of international scientific communication at that time:

Moro does not cite the works of Hooke and Ray, and although so many of his views were in accordance with theirs, he was probably ignorant of their writings, for they had not been translated. As he always refers to the Latin edition of Burnet, and a French translation of Woodward, we may presume

that he **did not** read English. (Lyell, 1830: 42).

A pilot study by David Banks (2005: 355) concluded that “Latin cannot be ignored as a contributing factor in the emerging scientific language towards the end of the 17th century.” Some new words combined Greek and Latin roots infelicitously, as Lyell (1833: 53) pointed out:

I have been much indebted to my friend, the Rev. W. Whewell, for assisting me in inventing and anglicising these terms, and I sincerely wish that the numerous foreign diphthongs, barbarous terminations, and Latin plurals, which have been so plentifully introduced of late years into our scientific language, had been avoided as successfully as they are by French naturalists, and as they were by the earlier English writers, when our language was more flexible than it is now. But while I commend the French for accommodating foreign terms to the structure of their own language, I must confess that no naturalists have been more unscholar-like in their mode of fabricating Greek derivatives and compounds, many of the latter being a bastard offspring of Greek and Latin.

## 1.2 The role of French, from 1066 onwards

After the Norman Conquest in 1066, Norman French became the language of the aristocracy, while Latin remained the language of religion, leaving English as the language of the common people. England and France coexisted as rival powers for many centuries, with much greater linguistic transfer from French to English. Several authors<sup>1</sup> have shown that the richness of English vocabulary stems from these three strands, with the persistence of quasi-synonyms derived from Latin, French, and English roots, e.g. secure, firm, and fast (Jackson & Amvela, 2007: 41).

In addition to Greek and Latin, academics were also expected to know vernacular languages; many scholars travelled extensively, as the following quotation shows:

During the 17th century, the German universities were of importance, along with the Dutch ones. Universities in England and France, and even in Italy and Spain, were also visited. The tradition of peregrination was upheld in the eighteenth century as well, though to a lesser degree. It is quite possible, though disputed, that the famous Swedish scholar Linnaeus, for example, bears witness to the learning of languages abroad; that his travelling pupils were versed in languages is well documented. (Gren-Eklund, 2011: 57)

---

<sup>1</sup> For a general overview, see Minkova & Stockwell, 2006.

The transition from Latin to the vernacular in scientific writing occurred progressively during the seventeenth century. Galileo wrote *Sidereus Nuncius* in Latin in 1610; his later works were written in Italian. Descartes wrote the French axiom “Je pense, donc je suis” in 1637, while his more famous “Cogito, ergo sum” appeared later, in 1644. Henry Oldenburg’s *Philosophical Transactions* (1665) often published texts translated from Latin or French, perhaps because the journal was “open to everyone” (Banks, 2009: 3). Newton wrote *Philosophiæ Naturalis Principia Mathematica* (1687) in Latin, while the later *Opticks* (1704) was written in English.<sup>2</sup> During the eighteenth century, “Saussure’s native French was by far the most important [language] in the world of the sciences.” (Rudwick, 2005: 29).<sup>3</sup>

Language learning may have been adversely affected by the French Revolution and the Napoleonic Wars that followed, preventing an entire generation of young men from undertaking their Grand Tour, previously considered an essential part of a gentleman’s education (Black, 2003). In 1795, Hutton included many French quotations in his *Theory of the Earth*, which was often judged “unreadable” by his peers (Leddra, 2010: 115). By 1830, Lyell would provide the general reader with translations and paraphrases. Dufief, the author of a French language manual, stated that “the English and Americans, when they begin the study of the French language, may be considered as *tabulae rasae*.” (1831: iv).

Long before the French Revolution (1789), D’Alembert pessimistically predicted:

Thus, before the end of the eighteenth century, a philosopher who would like to educate himself thoroughly concerning the discoveries of his predecessors will be required to burden his memory with seven or eight different languages. And after having consumed the most precious time of his life in learning them, he will die before beginning to educate himself. The use of the Latin language, which we have shown to be ridiculous in matters of taste, is of the greatest service in works of philosophy, whose merit is entirely determined by clarity and precision, and which urgently require a universal and conventional language. It is therefore to be hoped that this usage will be re-established, yet we have no grounds to hope for it. (1751, translated by Schwab, 1963: 92–93)

D’Alembert’s linguistic hopes were to be renewed after “the war to end all wars” (World War I, 1914–1918), with the quest for a new, more neutral language of peace.

<sup>2</sup> *Optice*, the Latin translation by Clarke, was published in 1706.

<sup>3</sup> Horace Bénédict de Saussure, the Swiss geologist (1740–1799).

### 1.3 The case for Esperanto

Ludwik Lejzer Zamenhof invented *Internacia lingvo* (International language) in 1887 “to unite the world through a common language” (Patterson & Huff, 1999: 444). *Doktoro Esperanto* (Dr. Hopeful) was in fact Zamenhof’s pseudonym (Gledhill, 2000: 4). In 1933, E. L. Thorndike estimated that Esperanto could be learnt in one-fifth of the time it took to learn French, German, Italian, or Spanish (Fettes, 1997: 152). Eugen Wüster, the terminologist, suggested that it would take six times more energy to learn English than to learn Esperanto (Wüster, 1934). Despite this ease of acquisition, conservative estimates indicate only 40,000 fluent speakers of Esperanto (Gledhill, 2000: 10).

Comparative figures for English, French and Esperanto are presented in Table 1, based on data collected by Sydney S. Culbert and published in the *World Almanac and Book of Facts* in 1994 (Comrie, 1996: 36–37). Lindstedt (2010: 69) estimated that “there are perhaps one thousand first-language speakers”.

Table 1: Number of 1<sup>st</sup> or 2<sup>nd</sup> language speakers

Language	1st language	2nd language	Total
English	322,000,000	148,000,000	470,000,000
French	72,000,000	54,000,000	124,000,000
Esperanto	1,000	2,000,000	2,001,000

In contrast with English and French, other major languages “are more restricted geographically, e. g. Spanish primarily to Spain and Latin America, Arabic to the Middle East and North Africa, Russian to the former USSR” (Comrie 1996: 37). The Universal Association of Esperanto has members in 120 countries (UEA, 2016). Google Maps and Chatterplot show Esperanto speakers in all time zones. Esperanto could therefore legitimately rank as a geographically widespread language, but with limited speaker density.

France is, surprisingly, the country spanning the greatest number of time zones (Table 2). “This unusual span is due to France’s scattered national territories. The areas in French Polynesia in the Pacific Ocean are mainly responsible for this.” (*India Today*, October 28, 2015). If the analysis were to be based on the number of time zones outside the mainland area, the United Kingdom would rank second.



Table 2: Number of Time Zones (Source: *India Today*, 2015)

Country	France	USA	Russia	UK	Australia
Mainland	1	6	11	1	3
Other	11	5	0	8	5
Total	12	11	11	9	8

In 1922, the League of Nations<sup>4</sup> proposed to facilitate the adoption of Esperanto as their working language by teaching it in schools, requiring ten votes in favour of the motion. Delegate Gabriel Hanotaux voted against the proposal, probably fearing that French would be endangered by the spread of Esperanto, and this single negative vote was sufficient for the motion to be rejected (Tafalla Plana, 2010: 139). A key opportunity to promote Esperanto as a major international language was lost, yet the international influence of French declined, as English gradually became the most widely used international language.

### 1.4 The ubiquity of English

Among the reasons for the spread of English are its capacity for integrating words from other languages, and the fact that there is no one variety of English with higher status than the others. “English is thus a language whose native speakers, especially those involved in international communication, are in general tolerant of a wide range of variation and recognize the existence of a number of non-native speaker varieties.” (Comrie, 1996: 49). Monolingualism may also explain this tolerance: in 2016, 65.4% of working-age adults in the UK reported that they knew no foreign languages, while the figure for France was only 39.9% (Eurostat, 2018).

One drawback of monolingualism is that it may limit access to scientific knowledge, and even distort knowledge of the history of science. Earth Science was often interpreted through the prism of locally accessible data, as the following reference to “British geology” indicates.

It might seem (to English-language readers) that Wernerism was decisively defeated by Hutton and British geology in general. However, while the rise of organized field-mapping, described in Chapter 5, was a notable expression of ‘English-language geology’, the greater part of the basic work in petrography was carried out on the Continent, particularly in France and Germany. (Oldroyd, 1996: 196).

<sup>4</sup> <http://vortareto.free.fr/argumentaire/sdn/sdn.pdf> (Accessed 3 February, 2019).

Translating scientific texts could provide a solution to the problem of monolingualism, but translation and stimulus meanings may not always coincide (Quine, 1960: 40). Wegener's "*Theorie der Kontinentalverschiebung*" was initially translated as the "theory of continental displacement", but it is much more widely known under the term "continental drift", despite or perhaps even because of its negative semantic prosody (Chateau, 2010). Problems may also arise from the general tendency towards lexical simplification in translated texts (Xiao, 2010: 29). In the absence of accurate, culturally relevant translations, the entire paradigm of normal science may become skewed (Kuhn, 1970).

An analysis of the languages cited as references in 139 sedimentological papers published by Elsevier in 1964 led to the following conclusions:

- (1) Russian and English were the two major scientific languages, but with little contact between the two areas, except for "a small proportion of translated literature".
- (2) French (or German) was seldom read outside France (or Germany).
- (3) British sedimentologists read more languages (especially German) than their North American colleagues, who tended to rely on "a smaller number of scientific periodicals".
- (4) In North America, older scientists read more languages than the younger generation.
- (5) Older Dutch sedimentologists read more French than the younger generation, while younger German sedimentologists read more languages than the older generation. (adapted from Mantel & van Dilst, 1965: 22-23)

In 1967, a similar survey by the British National Lending Library concluded that:

The foreign-language barrier facing British scientists is a function of a) the amount and value of foreign scientific literature produced; b) the linguistic ability of the people wishing to read it; and c) the availability and effectiveness of translation services." (Wood, 1967).

Barely forty years later, the barrier to understanding "foreign scientific literature" has almost disappeared, but not through greater knowledge of other languages. In the twenty-first century, the majority of the scientific literature in many fields tends to be published directly in English, which has obliged many professional translators to adapt to this situation, often by providing editing rather than translation services.

Although English, French, and German were used with similar frequency to write scientific articles at the end of the nineteenth century, French has been in constant decline since that time, according to data

collated by Ammon (2012: 338). The use of German increased quite rapidly until the aftermath of the First World War, even overtaking English, but has declined steadily since 1920. The peak in the use of Russian (from 1950 to 1988) is probably linked to geopolitical events, such as the Cold War, the founding of Solidarność, and Gorbachev's dual programme of Perestroika and Glasnost.

Many academics deplore this decline in multilingualism, yet it is undoubtedly true that the effort to master one foreign language is far less than that required for the mastery of the seven or eight languages that D'Alembert thought would be necessary should all scientists use the vernacular for their publications. As Marcia McNutt (2013: 13), Editor-in-Chief of *Science* pointed out: "Even the most brilliant scientific discovery, if not communicated widely and accurately, is of little value." Although it has often been an uphill struggle, at times of paradigm shift, for the authors of controversial theories to gain recognition by the scientific community, as was the case for Wegener and continental drift (Chateau, 2014), it is undoubtedly necessary today for scientists to communicate their discoveries, and a recent report by the American Sociological Association has gone so far as to suggest that making scientific discoveries known to the general public is even more important than communication within the scientific community (ASA, 2016).

## 2. Nurture: Advice to authors

In order to communicate both widely and accurately, some degree of mastery is necessary. In the field of scientific communication, the peer-review system is seen as a guarantee of quality. In the field of communication in English, the educated native speaker is considered to provide a model worthy of imitation. This can sometimes lead to problems of unwitting plagiarism.

This section explores the notion of native speaker, and then introduces a pilot corpus of Advice to Authors, to draw some conclusions about the constraints that apply to the concept of English as an international language for written scientific communication, which is not generally shown the tolerance accorded to oral communication (Ammon, 2012: 350).

### 2.1 Defining the model

In order to imitate the native speaker model, it is necessary to define the notion of native speaker. Swales (1985) proposed a series of criteria to differentiate between "Native Speaker (NS)" and "Non-Native Speaker

(NNS)” status, with plus or minus scores of up to 12 points (which would perhaps be judged less than politically correct by today’s standards). The terms proposed in this section are: “NES” (Native English Speaker), NFS (Native French Speaker), and NNES (Non-Native English Speaker).

1. **R** (1 point) Last name (including Anglicization etc.),
2. **RRR** (3 points) Institutional affiliation (assumed to be permanent unless otherwise indicated),
3. **R** (1 point) All citations to English language publications,
4. **RR** (2 points) First name of author (John, Jean, Juan, Giovanni, Johann, etc.; surprisingly useful evidence when available, especially useful for identifying the language status of married women),
5. **RR** (2 points) All self citations to English language publications,
6. **RRR** (3 points) NS/NSS Status-relevant footnotes or endnotes (Translation; acknowledgements for linguistic help; visiting affiliation and being on leave away from permanent base; we also think that source of funding falls into this category especially if the grant is being provided by an agency such as the Ministry of Defence). (Swales, 1985)

According to these criteria, Carmela Chateau Smith would have a minus score, ranking E (-5 to -12) on a scale from A (definitely NES) to E (definitely NNES). Far from invalidating Swales’ somewhat ethnocentric criteria, this result raises several interesting questions. Do birth and education confer permanent NS status? Does the acquisition of additional languages (French, in the present case) actually modify linguistic ability and status? Does bilingualism lead to a higher threshold of tolerance for variation?

The less stringent criteria proposed by David Banks might produce more informative and useful results.

The articles were selected on the basis of two criteria:

- (1) the article was written in English;
- (2) at least one of the authors provided an address in an English-speaking country.

The object of this selection was to isolate those articles which were written by English speakers, so as to avoid possible interference from other languages. (Banks, 1991: 59)

It is possible that the presence of a co-author residing in an English-speaking country may help to avoid cross-linguistic contamination, but many factors enter into play in the publication of a scientific article, from the Instructions or Guidelines to Authors, through the process of peer-review (potentially by NNES), to the final stages of copy-editing.

Furthermore, most articles in the “hard sciences” (Weijen, 2012) are not the work of a single author, but are the fruit of teamwork, with several co-authors.

Many medical journals now ask the corresponding author to identify the roles and contribution of any co-authors, with some version of the following four criteria, as proposed by the International Committee of Medical Journal Editors:<sup>5</sup>

Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND  
 Drafting the work or revising it critically for important intellectual content; AND  
 Final approval of the version to be published; AND  
 Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Unfortunately, none of the above criteria provides any information about the linguistic status of the contributors, and traces of copy-editing are difficult to identify, unless various versions of the article are available, as in the case of self-archiving of a pre-print copy, for example.

## 2.2 The Advice for Authors corpus

In order to explore the recommendations and advice available to researchers, a pilot Advice for Authors corpus was created, using *Sketch Engine* (Kilgarriff et al., 2014).<sup>6</sup> This corpus is composed of 27 articles providing advice to authors on how to write for publication, in the field of medical research; the total word count is 97,992 (or 119,185 tokens, including punctuation); the longest article contains 17,660 tokens, and the shortest contains 918 tokens. The corpus could, in fact, be considered to contain 20 articles (with an average word count of 6,000 tokens), since the shortest texts form eight parts of a series, with the same authors for each text. The articles were published over a ten-year period, between 2006 and 2015. Restricting the topic to a single field eliminates domain-specific variability, but no further precautions were taken to limit the range and scope of the language used. The authors (whether NES or NNES) are from

---

<sup>5</sup> Recommendations available online: (Accessed 3 February 2019)  
 <<http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html#two>>.

<sup>6</sup> Online corpus software and web service, with access to pre-loaded corpora:  
 <<http://www.sketchengine.co.uk/>>.

several different countries, of various scientific backgrounds, and offer advice on all stages of publication. This pilot corpus is therefore presented as a valid, random sample of the advice available to authors in various sub-fields of medical research. Brief details of the texts in the Advice corpus are given in Appendix I.

### 2.3 Exploring the corpus

*Sketch Engine* can be used to draw up word lists. It is possible to create word, lemma, or even lemmatized parts-of-speech lists, with an additional lowercase option. Unsurprisingly, in the basic frequency list for the Advice corpus (with the lowercase option), the fifteen most frequent words all belong to the closed set class, and are very short, confirming the Zipfian principle that: “as the relative frequency of a word increases, it tends to diminish in magnitude.” (Zipf, 1935: 38). Among the first twenty words in order of frequency, only two are more than four letters in length, and both are lexical in nature, from the open-set class. These two words together, “journal” and “writing”, encapsulate the main theme of the Advice corpus. The most frequent collocation for the lemma “journal/s” is not “writing” (22 occurrences), but “published” (43 occurrences). The implication is therefore that being published is twice as important as writing.

Writing for publication tends to be a choral work rather than a solo performance, even when an article has only one named author. In the Advice corpus, as expected, the plural forms, “authors” and “reviewers”, found in 21 texts, are more frequent than the singular “author” and “reviewer”. The word “review” occurs in 23 texts, underlining the importance of this task (Table 3).

Table 3: Advice for author.\*, peer.\*, and review.\*

	T	F		T	F		T	F		T	F
author	25	136	peer	14	58	review	23	249	reviewer	16	87
authors	21	244	peer-review	5	5	reviews	10	65	reviewers	21	259
authorship	9	33	peer-reviewed	10	35	reviewed	12	26	reviewing	11	55

T= number of texts; F= number of occurrences

Focusing on the most frequent words common to all 27 texts in the Advice corpus can help to identify the main preoccupations of their authors, but most of the 29 results are short, closed-set words, with only two nouns: “journal” and “paper”, again underlining the importance of publication.

The two most frequent words, consistent with most corpus results, are “the” and “of”, exemplifying the most frequent frame in English: “the” \*

“of”\*, which produces a concordance of 740 occurrences, the most frequent of which are presented in Table 4, classified by frequency and by category of use.

Table 4: The most frequent frames and their uses.

Size and quality		Results		Options		Affiliation	
the quality of	40	the outcome of	12	the use of	11	the journal of	10
the number of	30	the results of	12	the choice of	6	the Departments of	8
the importance of	20	the end of	11	the style of	6	the University of	8
the lack of	8	the findings of	6	the type of	6	the work of	6
the accuracy of	7	the reporting of	6	the process of	5	the editor of	5
				the subject of	5	the names of	5
<b>Semantics</b>		<b>Probability</b>				the role of	5
the purpose of	6	the chances of	5				
the meaning of	5	the likelihood of	5				

The most frequent verb forms are, in order: is, be, are; have; should, will, may. This result provides useful information for NNES. It is possible to create concordances in *Sketch Engine* using Corpus Query Language (CQL). The following expression: [tag="VB.\*"] [tag="VVN.\*"] produces a list of 1157 uses of passive forms (any form of the verb “be”, followed by any past participle). The large number of occurrences of “be” (429) suggests a further query, using the expression: [tag="MD.\*"] [tag="VB.\*"] [tag="VVN.\*"] to find modal passive forms (250), presented in order of frequency: should (89), can (50), may (33), will (33), must (22), could (12), would (6), and might (5). The use of “should” is perfectly natural in a corpus of advice, but the proportion of “can” and “may” in relation to “must” suggests that few hard and fast rules can be proposed.

To examine this hypothesis, three CQL expressions<sup>7</sup> were used to identify the deliberate use of the “split infinitive”, a well-known example of the prescriptive versus the descriptive approach to grammar (Yule, 2010: 85–87). The results (43 split infinitives, in 18 texts) indicate that splitting the infinitive is not a problem for most Advice corpus authors. The expression “adverb + to + verb” (66 occurrences) was used only once to deliberately avoid the split infinitive (“Do not be afraid respectfully to argue your case”). There were only eight cases where the expression “to + verb + adverb” (75 occurrences) could potentially have been replaced by the split infinitive.

<sup>7</sup> 1° "to" [tag="RB.\*"] [tag="V.\*"], 2° [tag="RB.\*"] "to" [tag="V.\*"], 3° "to" [tag="V.\*"] [tag="RB.\*"].



The final example from the Advice corpus is “English” (see concordance data in Appendix II and Appendix III<sup>8</sup>), with 12 occurrences of English as an adjective, and 21 as a noun. The word “English” is used in only 8 of the 27 articles in the corpus (texts T1, T3, T11, T13–16, and T24) suggesting a subset with a more specific focus on language. Closer examination indicates that the focus in T11 is American English, while T15 makes several references to British English, but also mentions German as an example of a foreign-language reference, echoing the third result from Manten & van Dilst’s study (section 1.4). The only language mentioned in collocation with “English” is Spanish (10 occurrences, all in T24). French is mentioned twice in T13, but as a nationality, not as a language. Taiwanese is referred to in T16 as the nationality of potential readers, to underline the importance of writing for an international audience.

Given the emphasis frequently placed on the importance of “Native-Speaker English”, or rather of having one’s work edited, revised, or even “vetted by a native speaker” (Carli & Ammon, 2007), it is rather surprising that two-thirds of the articles in the Advice corpus make no mention whatsoever of “English”. It is possible that, in the twenty-first century, it is no longer considered necessary to labour the language point. Some bias may also be found in the fact that all the articles were published in English. Perhaps articles in other languages might provide more advice for NNES on how best to be published in English.

### 3. Norms: Reviewers, recommendations and remarks

In 1999, a survey by David Banks indicated that 70% of respondents (French researchers) considered that they were ill-prepared to speak in English at conferences.

Taking part in spontaneous debate and fielding questions is evidently an area where French scientists feel insufficiently confident, and consequently at a disadvantage in the international situation.” (Banks, 1999: 6).

Great effort has been spent since that time on developing language courses at French universities for students of other disciplines, and yet, despite the success of such endeavours, the moment of truth for many scientists whose first language is not English still occurs at question time.

This study focuses on another problematic aspect of international communication, that of writing in English in order to achieve an acceptable standard for publication. Most researchers whose native language is not

---

<sup>8</sup> The Word Sketch of “English” is presented in Appendix III.



English will have at some point received a review with the potentially damning comment: “Should be reread by a native speaker”. Carli and Ammon (2007) have explored in some detail the continuing problem of linguistic inequality, but there is still little quantitative assessment of the time spent by scientists rewriting papers to satisfy native-speaker norms.

This final section presents a small case study, from an *ad hoc* micro-corpus of documents, including replies from the editor, reviewers’ remarks, and copy-editing documents, collected in the early twenty-first century. Problems of several types are encountered, not always necessarily linked to language.

### 3.1 Publish or perish in English?

Although English is the main language for international scientific communication, there are still fields where publication in other languages is of equal importance. The choice of language may be related to the type of science, with arts and humanities on one side and the physical sciences on the other. Table 5 (adapted from Weijen, 2012) presents the ratio of publications in English compared to native-language publications, and the relative importance of native-language publications for the four main categories of science, based on data from Scopus for the period 1996–2011. The first column indicates the ratio of publications in English for a given country, compared to those published in the native language of the country. Thus, in Germany, for every article published in German, nine articles are published in English. Distribution figures for English-language publications show that nearly 45% of publications are in the physical sciences, almost 25% in the life sciences, nearly 20% in the field of health, but only 10% in that of the arts. Within Europe, combined figures for native-language publications in the “soft” sciences range from 56% for German, with a fairly low ratio of publications in English, to 79.2% for Italian, with a fairly high ratio of publications in English. In contrast, Dutch has the highest ratio of publications in English and yet the second highest figure (78.4%) for native-language publications in the “soft sciences”.

Table 5: Ratio of English & relative % by field of publication, 1996–2011  
(after Weijen, 2012)

English ratio 1996–2011	Language	Hard Sciences		Soft Sciences		Other	Total
		Life	Physical	Health	Arts		
	English	23.4	44.7	19.5	10.7	1.7	100%
9:1	German	7.3	34.5	32.5	23.5	2.2	100%
43:1	Dutch	14.9	3.2	52.3	26.1	3.5	100%
7:1	French	8.6	16.3	36.4	36.5	2.3	100%
30:1	Italian	4.7	12.1	38.6	40.6	4.0	100%
6:1	Spanish	10.8	13.2	44.4	29.6	2.0	100%
3:1	Portuguese	26.1	11.5	38.4	22.1	1.9	100%
27:1	Russian	17.2	45.0	21.0	8.4	8.4	100%
2:1	Chinese	8.7	72.5	13.0	2.9	2.9	100%

Two factors could possibly explain why the author's native language is chosen in preference to international English, particularly in the "soft" sciences. Since the hard sciences tend to be more factually based, a larger proportion of the information can, and indeed should, be conveyed in the form of graphs, tables and other figures, thus lessening the linguistic pressure on the author. Articles in the humanities often rely more on words than on graphs and figures. The range of lexis involved may therefore be broader, particularly as a picture is worth a thousand words. The type/token ratio (TTR), where the number of words in a given document is divided by the number of different words used in the document, is used in corpus linguistics to quantify lexical density. A scientific article, although using very specific terms, often infrequent in general English, will tend to use the same words more often, while a more literary article will tend to use a broader range of synonyms to make the text less monotonous to read. Work on different types of texts has confirmed that "the specialised corpus highlights a small, probably technical vocabulary" (Sinclair, 2004).

Getting a text accepted for publication requires a great deal of work, and the additional effort required when writing in a foreign language is not necessarily rewarded. The final goal for researchers today is not merely publication, but publication followed by citation. Science truly becomes science when it has been read, used, and quoted by other scientists. Journals base their impact on citation figures, and yet, even today, specific language networks still exist within the broader scientific community, particularly when fieldwork is locally based, as in archaeology, where publication in the language of the country (e.g. French, German, Italian, etc.) is still perfectly acceptable. Certain studies in climatology, for example, may also be preferentially published in French (e.g. *Climatologie*. Journal de l'Association Internationale de Climatologie). If scientists are communicating in two or

more languages, the focus on publication in English may be less strong, but the effort required may be even greater.

### 3.2. Publish and flourish in Spanish?

The following example is taken from a study published in Spanish, accompanied by an English version of the Spanish abstract. No translator is identified or otherwise acknowledged in the article. The English abstract contains several linguistic errors, as do the Google Translate versions (produced three years apart). All three translations are presented here, below the original Spanish. The French reader may find the Spanish original easier to understand than the English translations.

**Resultados:** España se sitúa en la decimoprimer posición mundial y en séptima posición dentro de sus homólogos europeos. La producción española supone el 2,4% del total de documentos biomédicos del mundo, aunque solamente el 1,8% de citas recibidas. Cerca de la mitad de todos los documentos van con firma del sector sanitario. El 24,8% de los documentos se firma en colaboración internacional, proporción que acumula el 45,5% de todas las citas. (Camí et al., 2005)

**Results:** Spain occupies the eleventh place in the world, and the seventh place among the European countries. Despite Spanish contribution to the world biomedical production is 2,4% it accounts for only 1,8% of the whole citations. Spanish health sector is responsible for nearly the half of the total biomedical output. The 24,8% of documents are published in international collaboration, a proportion that accumulates the 45,5% of biomedical citations received. (Camí et al., 2005)

**Results:** Spain is the world's eleventh and seventh position within their European counterparts. The Spanish production represents 2.4% of total world biomedical documents, although only 1.8% of citations received. About half of all documents go with signing the health sector. 24.8% of the documents signed at international collaboration, a proportion that accumulates 45.5% of all appointments. (Google Translate: April 2016)

**Results:** Spain is in the eleventh position in the world and seventh in its European counterparts. Spanish production accounts for 2.4% of all biomedical documents in the world, although only 1.8% of citations received. Nearly half of all documents are signed by the health sector. 24.8% of the documents are signed in international collaboration, a proportion that accumulates 45.5% of all citations. (Google Translate: February 2019)

Would the text have been as easy to find if no English abstract had been proposed? Would the Google translation have been sufficient to attract readers? Is it legitimate to include an article in one's list of references when only the abstract has been read and fully understood? More importantly, is

the fact that Spanish production represents 2.4% of the global biomedical publication, but only 1.8% of citations a statistically significant result? The implication is that, if publishing in English does not result in citation, scientists might just as well publish in their native language, which requires far less initial effort. As Françoise Salager-Meyer points out:

when no editorial support services are available [...] a solution for EAL<sup>9</sup> writers in such contexts would be to write their papers in their own language and publish them in top-quality journals, i.e. journals indexed in international databases. (Salager-Meyer, 2009: 47).

### 3.3 Under Review: a case study

It is a truth universally acknowledged that a language teacher in a science faculty, in possession of NES status, must be in want of Native French-Speaking researchers (NFS) seeking to publish in English. Reviews and papers contributed by colleagues in Earth & Environmental Sciences form the Review collection.

This case study is based on an article finally published in the print version of the *International Journal of Climatology* in July 2015. The publication timeline provides a rough estimate of the time scientists spend rewriting in English, once all the scientific work has been completed. The article was published online on 16 September 2014. It was accepted for publication on 11 August 2014, fairly soon after the revised version was submitted, on 24 July 2014, but the initial version of the article had been submitted on 11 October 2013. This elephantine gestation period, from computer to paper, includes two rounds of reviewing, and therefore three versions of the paper, but the main focus of this case study is the English revision phase, covering a period of two weeks (July 10–24, 2014).

This revision period began with the following call for help from a colleague, after a paper had been granted “conditional acceptance”, subject to “satisfactory revisions”. The frustrating phrase is, of course, “a native English writer”.

La phrase qui me frustre le plus est de nouveau tombée, pour le papier d'un ancien thésard du labo (en copie)<sup>10</sup>: “I would follow the suggestion of the reviewer 1 and suggest a careful review of a native English writer.”  
Arghhhhhhh..... ! (Benjamin Pohl, 25 June 2014, personal communication)

<sup>9</sup> English as an Additional Language (EAL).

<sup>10</sup> The sentence I find the most frustrating has been pronounced, yet again, for a paper by a former PhD student of the lab (cc'd):

The sentence by Reviewer 2 (with its somewhat unfortunate errors, in bold) echoes an equally frustrating remark by Reviewer 1, from the first review stage:

Finally, there are numerous places where the writing is awkward. [...] There are many other examples. I suggest the authors request a careful review of the writing by a native English writer.

The difference between “by” and “of” is perhaps masked in other languages, where the same expression might be used to express both “a review of Shakespeare’s plays” and “a play by Shakespeare”. This type of error can sometimes produce amusing results, but few readers would misinterpret Reviewer 2’s meaning, given the context of the remark. Both reviewers gave useful linguistic advice to the authors in the first round of revision. Reviewer 1 (probably a “native English writer”) provided four examples of “awkward” writing, while Reviewer 2 provided help in the form of at least 30 line-referenced remarks, correcting typos (conems -> concern), suggesting vocabulary (replace: map -> show), and identifying missing words (missing: 5352... word missing). One remark, however, led to a breakdown in communication: “change: cluster for whose pattern the Euclidian distance can be minimized” was answered with the remark: “We are not quite clear by what you mean, so we did not add this suggestion.” (J.C., 14 May 2014, personal communication)

Communication breakdowns may sometimes lead to far more disastrous consequences, as the following example demonstrates:

On January 25, 1990, an Avianca B-707 crashed during an approach to the John F. Kennedy Airport in New York. [...] This accident, like that at Tenerife years before, demonstrated the weakness of voice communications that take place in other than one’s native language. Many factors were present in this accident as in the Tenerife one, but basic was the inability of the crew to communicate effectively with the controller. Standardized, understandable voice phraseology was adopted by the FAA subsequent to this accident. (Wickens et al., 1997: 29).

Avoiding this type of ambiguity is one of the science writer’s main responsibilities, in order to communicate “widely and accurately” (McNutt, 2013: 13), although the failure to communicate is rarely as deadly as in this tragic example.

Should EAL reviewers also be required to seek linguistic assistance when writing their reviews? Are their remarks intended to serve as an example to follow? Colleagues do occasionally ask for help, particularly when negatively reviewing a paper, in order to be sure that their remarks



will be correctly interpreted. But linguistic acceptability is a sliding scale, as the following quotations show:

Since both guest editors share the same views, the contributions to the present volume were not always vetted by native speakers, as long as verbal expressions appeared clear enough to the issue editors and acceptable to the reviewers. (Carli & Ammon, 2007).

The reviewer's reservations about publications in English that are not vetted by native speakers are vindicated by the current volume. (Hellmann-Rajanayagam, 2010).

One caveat is that “vetting” is a fairly negative process, linked to gatekeepers, and other authority figures. The second is that not all “native speakers” are competent to act as judges of scientific language, while many non-native speakers of English are perfectly able to produce correct English, without any assistance (except that of an editor), as the following example confirms:

*Pour ce qui est de mes articles et livres en anglais, je les ai écrits seul en anglais.* (I wrote my articles and books in English, unaided. Xavier Le Pichon, 3 April 2009, personal communication).

Linguistic ability is not the only factor in getting published. Certain aspects of language can be a challenge even for native speakers of English, particularly at a moment of paradigm shift. The following remark shows how important the choice of words can be, and how much reviewers can adversely affect the progress of science when a paper does not correspond to their notion of what falls within the realm of “normal science” (Kuhn, 1970).

I know that Harry Hess was very conscious of the conservative nature of geologists at the time and their aversion to speculation. Hence he tread very carefully. I too was very careful, for example, not to mention ‘continental drift’ explicitly.” (Vine, 10 January 2006, personal communication).

Vine's most famous text, cited 800 times<sup>11</sup>, contains the following discreetly cloaked references to “theory” and “drift”, within the same paragraph, two sentences apart.

The theory is consistent with, in fact virtually a corollary of, current ideas on ocean floor spreading<sup>7</sup> and periodic reversals in the Earth's magnetic

---

<sup>11</sup> According to *CrossRef*, *Google Scholar* indicates 1554 citations.

field<sup>5</sup>. [.../...] Thus, if spreading of the ocean floor occurs, blocks of alternatively normal and reversely magnetized material would drift away from the centre of the ridge and parallel to the crest of it. (Vine & Matthews, 1963: 948). *N.B. The numerals in superscript served to identify references.*

A mere five years later, the geological paradigm had shifted to such an extent that Vine could confidently refer to “continental” and “drift”, both separately and together, three times in the first five lines of his article, while the “if” and “would” of 1965 have become the “extremely probable” of 1968 (cited only 18 times).

On the basis of the evidence from continental geology, such as has been presented in the previous papers in this symposium, it is extremely probable that continental drift has occurred. It seems a logical step therefore to turn to the ocean basins, to the results of marine geology and geophysics, for some record of drift. The evidence for continental drift from submarine geology comes in the form of the evidence for sea-floor spreading. The concept of sea-floor spreading is a comparatively recent one. It was revived and first formulated in some detail by Hess (1962). (Vine, 1968: 325).

## Conclusion

The time spent on collaborative language revision may be fruitful for both the language teacher and the scientist, and should probably be integrated into the workload of university language teachers. The most successful collaborations occur when the experienced professor agrees to work with the language teacher and the junior researcher. The professor knows the field of study, the young researcher learns how to write for publication, and both benefit from the assistance of the language teacher, who acquires valuable information about a specific language community. Team co-authoring may produce successful results, as the following reference shows:

Crétat, Julien, Benjamin Pohl, Carmela Chateau Smith, Nicolas Vigaud & Yves Richard. 2015. An original way to evaluate daily rainfall variability simulated by a regional climate model: the case of South African austral summer rainfall. *International Journal of Climatology*. 35(9). 2485–2502.

This study has focused on Earth and Environmental Science (often negatively described as “hard”), but other fields (the equally pejorative “soft” sciences) should not be forgotten.

Evaluation of research in the Social Sciences and Humanities (SSH) is a thorny problem as the bibliometric measures designed for the life sciences simply do not apply. (Galleron & Williams, 2013: 102).

An article entitled *Scientific utopia* suggests ways of improving the situation, which could help to level the playing field for non-native writers of English, and might also help colleagues in fields less commercially interesting than Medicine or Earth Science.

We call for six changes: (1) full embrace of digital communication, (2) open access to all published research, (3) disentangling publication from evaluation, (4) breaking the “one article, one journal” model with a grading system for evaluation and diversified dissemination outlets, (5) publishing peer review, and (6) allowing open, continuous peer review. (Nosek & Bar-Anan, 2012).

Emulating expert behaviour is often recommended as the best path to success. Xavier Le Pichon, sole author of “Sea-floor spreading and continental drift” (1968; 767 citations to date), and lead author of *Plate Tectonics* (1973), a book written directly in English, gave the following useful, but very pragmatic replies:

CCS: Avez-vous un secret, ou est-ce simplement le fruit d'un long travail?  
 XLP: Aucun. Anglais scolaire avec un été en Angleterre à 11 ans et un aux US à 15 ans puis un an aux USA en 59 et 5 ans entre 63 et 68.  
 CCS: Auriez-vous des ressources bibliographiques à conseiller à mes étudiants pour la rédaction en anglais?  
 XLP: Aucun sinon des séjours en milieu anglophone. La plupart de mes bons étudiants écrivent un bon anglais scientifique, quelquefois meilleur que leur français!

English version (by CCS).

Do you have a secret, or is it simply the result of time and hard work?  
 No secret. School English, with a summer in England at the age of 11, and one in the US at 15, then one year in the USA in 1959, and five years between 1963 and 1968.  
 Can you recommend any resource materials for my students about writing in English?  
 None, apart from language stays in English-speaking countries. Most of my good students write good scientific English, sometimes better than their French!” (Xavier Le Pichon, 3 April 2009, personal communication)

Exchange programmes combining language and science (e.g. Erasmus) may therefore provide part of the answer to the thorny problem of international scientific communication.



## References

- D'Alembert, Jean Le Rond. 1963. Preliminary Discourse to the Encyclopedia of Diderot. Translated by Richard N. Schwab. Chicago: University of Chicago Press.
- American Sociological Association (ASA). 2016. What Counts? Evaluating Public Communication in Tenure and Promotion. Final Report of the American Sociological Association Subcommittee on the Evaluation of Social Media and Public Communication in Sociology. [http://www.asanet.org/sites/default/files/tf\\_report\\_what\\_counts\\_evaluating\\_public\\_communication\\_in\\_tenure\\_and\\_promotion\\_final\\_august\\_2016.pdf](http://www.asanet.org/sites/default/files/tf_report_what_counts_evaluating_public_communication_in_tenure_and_promotion_final_august_2016.pdf) (Accessed 3 February 2019)
- Ammon, Ulrich. 2012. Linguistic inequality and its effects on participation in scientific discourse and on global knowledge accumulation – With a closer look at the problems of the second-rank language communities. *Applied Linguistics Review*. 3(2). 333–355.
- Banks, David. 1991. Some observations concerning transitivity and modality in scientific writing. *Language Sciences*. 13(1). 59–78.
- Banks, David. 1999. Becoming part of the network: French scientists and the use of English at conferences. *ASp*. 23–26. <<http://asp.revues.org/2442>>.
- Banks, David. 2004. Philosophy, science, ideology and the establishment of science as a profession. *ASp*. 43–44. <<http://asp.revues.org/1084>>.
- Banks, David. 2005. On the historical origins of nominalized process in scientific text. *English for Specific Purposes*. 24(3). 347–357.
- Banks, David. 2009. Creating a specialized discourse: the case of the *Philosophical Transactions*. *ASp*. 56. DOI : 10.4000/asp.98
- Black, Jeremy. 2003. Italy and the grand tour. New Haven and London: Yale University Press.
- Camí, Jordi, Ey Suñén-Piñol, and Raúl Mendez-Vasquez. 2005. Mapa bibliométrico de España 1994-2002: biomedicina y ciencias de la salud. Bibliometric map of Spain 1994-2002: biomedicine and health sciences. (Abstract in English). *Medicina clínica*. 124(3). 93–101.
- Carli, Augusto and Ulrich Ammon. 2007. Linguistic inequality in scientific communication today. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chateau, Carmela. 2010. Suivre la dérive de “continents” dans un corpus diachronique d’anglais géologique. In Banks, David (ed.). *Aspects diachroniques du texte de spécialité*. Paris: L’Harmattan. 21–34.
- Chateau, Carmela. 2014. Did you feel the Earth move? Examples of humour in the geological sciences. In Coloru, Omar and Giuseppe Minnuno

- (eds.). CD-ROM. Crosscultural humour: Images, aspects and languages. 181–191.
- Comrie, Bernard. 1996. Natural and artificial international languages: A typologist's assessment. *Journal of Universal Language*. 3. 35–55.
- Diderot, Denis and Jean Le Rond D'Alembert. 1751. *Encyclopédie ou dictionnaire raisonné des sciences, des arts et des métiers*, Volume 1. Paris: Briasson.
- Dufief, Nicolas Gouin. 1831. Nature displayed in her mode of teaching language to man. London: J. and C. Adlard.
- Eurostat, 2018. (Accessed 3 February 2019)  
[https://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign\\_language\\_skills\\_statistics#Number\\_of\\_foreign\\_languages\\_known](https://ec.europa.eu/eurostat/statistics-explained/index.php/Foreign_language_skills_statistics#Number_of_foreign_languages_known)
- Fettes, Mark. 1997. Esperanto and Language Awareness. *Encyclopedia of Language and Education*. 6. 151–159.
- Galleron, Ioana and Geoffrey Williams. 2013. The good, the bad and the downright mediocre: Quality judgments in book reviews. *Word and Text: A Journal of Literary Studies and Linguistics* 3(1). 102–118.
- Gledhill, Christopher. 2000. The grammar of Esperanto. A corpus-based description. Munich: Lincom Europa.
- Green, Tamara M. 2015. The Greek and Latin roots of English. 5th ed. London: Rowman and Littlefield.
- Gren-Eklund, Gunilla. 2011. Philology in the eighteenth century: Europe and Sweden. In Gunnarsson, Britt-Louise (ed.). *Languages of science in the eighteenth century*. Berlin: Walter de Gruyter. 45–61.
- Hellmann-Rajanayagam, Dagmar. 2010. Review of Kazuki Iwanaga (ed.): *Women's political participation and representation in Asia. Obstacles and challenges*. *Journal of Current Southeast Asian Affairs*. 29(1). 157–162.
- Horrocks, Geoffrey. 2010. *Greek. A history of the language and its speakers*. 2nd ed. London: Wiley-Blackwell.
- India Today. 2015. The world is divided by time: List of countries with most number of time zones. October 28, 2015.  
<http://indiatoday.intoday.in/education/story/time-zones/1/509120.html>.
- Jackson, Howard and Etienne Zé Amvela. 2007. *Words, meaning and vocabulary: an introduction to modern English lexicology*. 2nd ed. London: Continuum.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography* 1(1). 7–36.
- Kuhn, Thomas S. 1970. *The Structure of Scientific Revolutions*. 2nd ed. Chicago: University of Chicago Press.

- Leddra, Michael. 2010. *Time matters: Geology's legacy to scientific thought*. London: Wiley-Blackwell.
- Lindstedt, Jouko. 2010. Esperanto as a family language. In Dervin, Fred (ed.). *Lingua francas: La véhicularité linguistique pour vivre, travailler et étudier*. Paris: L'Harmattan. 69–80.
- Lyell, Charles. 1830. *Principles of Geology*. Volume I. London: John Murray.
- Lyell, Charles. 1833. *Principles of Geology*. Volume III. London: John Murray.
- Manten, A.A. and F. J. H. van Dilst. 1965. The problem of language in Earth Sciences. *Atlas (Supplement to Elsevier Journal of Geo-sciences)*. 1/2. 22–23.
- McNutt, Marcia. 2013. Improving scientific communication. *Science*. 342(6154). 13–13.
- Minkova, Donka and Robert Stockwell. 2006. 20 English Words. *The handbook of English linguistics*. 461–482.
- Mortensen, Janus and Hartmut Haberland. 2012. English—the new Latin of academia? Danish universities as a case. *International Journal of the Sociology of Language*. 216. 175–197.
- Nosek, Brian A. and Yoav Bar-Anan. 2012. Scientific utopia: I. Opening scientific communication. *Psychological Inquiry* 23(3). 217–243.
- Oldroyd, David. 1996. *Thinking about the Earth: A History of Ideas in Geology*. Cambridge, MA: Harvard University Press.
- Patterson, Robert and Stanley M. Huff. 1999. The decline and fall of Esperanto: Lessons for Standards Committees. *Journal of the American Medical Informatics Association*. 6(6). 444–446.
- Quine, Willard Van Orman. 1960. Translation and meaning. In *Word and Object*. Cambridge, MA: M. I. T. Press. 26–79.
- Ross, Sydney. 1962. Scientist: The story of a word, *Annals of Science*, 18:2, 65-85, DOI: 10.1080/00033796200202722
- Rudwick, Martin. 2005. *Bursting the limits of time: The reconstruction of geohistory in the age of revolution*. Chicago: University of Chicago Press.
- Salager-Meyer, Françoise. 2009. English as Tyrannosaurus rex or linguistic diversity? *The Journal of the European Medical Writers Association*. 18(1). 47–49.
- Sinclair, John McHardy. 2004. Corpus and text — Basic principles. In Wynne, Martin (ed.). *Developing linguistic corpora: a guide to good practice*. Oxford: Oxbow Books: 1–16.

- Société des Nations. (League of Nations). 1922. L'Espéranto comme langue auxiliaire internationale. Rapport du Secrétariat general adopté par la troisième Assemblée de la Société des Nations.  
<<http://vortareto.free.fr/argumentaire/sdn/sdn.pdf>>.  
(Accessed 3 February 2019)
- Stockwell, Robert and Donka Minkova. 2001. English words: History and structure. Cambridge: Cambridge University Press.
- Swales, John M. 1985. English language papers and authors' first language: Preliminary explorations. *Scientometrics* 8.1-2. 91-101.
- Tafalla Plana, Montserrat. 2010. El multilingüismo en la organización de las naciones unidas Revista de Llengua i Dret. 53. 137-162
- UEA. Universala Esperanto-Asocio. (Universal Esperanto Association). 2019. An update on Esperanto.  
<[https://uea.org/info/en/ghisdate\\_pri\\_esperanto](https://uea.org/info/en/ghisdate_pri_esperanto)>.  
(Accessed 3 February 2019)
- Vine, Frederick J. 1968. Evidence from submarine geology. Proceedings of the American Philosophical Society. 112(5). 325-334.
- Vine, Fred J. and Drummond Hoyle Matthews. 1963. Magnetic anomalies over oceanic ridges. *Nature*. 199(4897). 947-949.
- Weijen, Daphne van. 2012. The Language of (Future) Scientific Communication. *Research Trends*. 31.  
<<http://www.researchtrends.com/issue-31-november-2012/the-language-of-future-scientific-communication/>>. (Accessed 3 February 2019)
- Wickens, Christopher D., Anne S. Mavor and James P. McGee. (eds.). 1997. Flight to the future: Human factors in air traffic control. Washington: National Academies Press.
- Wood, David N. 1967. The Foreign-Language Problem Facing Scientists and Technologists in the United Kingdom—Report of A Recent Survey. *Journal of Documentation*. 23(2). 117-130.
- Wüster, Eugen. 1934. Grundzüge der Sprachnormung. Berlin: VDI-Verlag.
- Xiao, Richard. 2010. How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistics*. 15(1). 5-35.
- Yule, George. 2010. The study of language. 4th ed. Cambridge: Cambridge University Press.
- Zipf, George Kingsley. 1935. The psychobiology of language: An introduction to dynamic philology. Boston: Houghton-Mifflin.

## **Appendix I: Advice texts (alphabetical order, title, year, journal)**

- 1 A-Clarity-Clinic-for-Surgical-Writing\_2008\_Journal-of-Surgical-Research
- 2 Designing-conducting-and-reporting-clinical-research-A-step-by-step-approach\_2006\_Injury
- 3 Effective-writing-and-publishing-of-scientific-papers-Introducing-a-new-series\_2013\_Journal-of-Clinical-Epidemiology
- 4 Effective-writing-and-publishing-scientific-papers-part-I-how-to-get-started\_2013\_Journal-of-Clinical-Epidemiology
- 5 Effective-writing-and-publishing-scientific-papers-part-II-title-and-abstract\_2013\_Journal-of-Clinical-Epidemiology
- 6 Effective-writing-and-publishing-scientific-papers-part-III-introduction\_2013\_Journal-of-Clinical-Epidemiology
- 7 Effective-writing-and-publishing-scientific-papers-part-VII-tables-and-figures\_2013\_Journal-of-Clinical-Epidemiology
- 8 Effective-writing-and-publishing-scientific-papers-part-IX-authorship\_2013\_Journal-of-Clinical-Epidemiology
- 9 Effective-writing-and-publishing-scientific-papers-part-X-choice-of-journal\_2014\_Journal-of-Clinical-Epidemiology
- 10 Effective-writing-and-publishing-scientific-papers-part-XII-responding-to-reviewers\_2014\_Journal-of-Clinical-Epidemiology
- 11 Enhancing-the-Mission-of-Academic-Surgery-by-Promoting-Scientific-Writing-Skills\_2007\_Journal-of-Surgical-Research
- 12 Facilitating-writing-for-publication\_2008\_Physiotherapy
- 13 Formulating-an-Answerable-Question-Displaying-Data-Illustrating-Writing-Reviewing-and-Editing-Manuscripts-for-Publication-in-Medical-Journals\_2012\_The-American-Journal-of-Cardiology
- 14 Getting-published-in-peer-reviewed-journals\_2011\_International-Journal-of-Oral-and-Maxillofacial-Surgery
- 15 How-do-you-get-your-paper-published-more-quickly-in-the-British-Journal-of-Oral-and-Maxillofacial-Surgery-\_2015\_British-Journal-of-Oral-and-Maxillofacial-Surgery
- 16 How-to-get-your-paper-accepted-for-publication\_2012\_Paediatric-Respiratory-Reviews
- 17 How-to-get-your-paper-published-paper-An-editor-s-perspective\_2014\_Journal-of-Diabetes-and-its-Complications
- 18 How-to-publish-a-scientific-manuscript-in-a-high-impact-journal\_2014\_Advances-in-Digestive-Medicine

- 19 How-to-write-and-publish-an-original-research-article\_2010\_American-Journal-of-Obstetrics-and-Gynecology
- 20 How-to-write-a-research-paper\_2012\_Paediatrics-and-Child-Health
- 21 How-to-write-a-scientific-article\_2010\_Osteopathic-Family-Physician
- 22 Improving-the-quality-of-papers-submitted-to-dental-journals-Transcription-of-session-for-editors-associate-editors-publishers-and-others-with-an-interest-in-scientific-publishing held at IADR meeting in Cape Town on Wednesday, 25 June 2014\_Journal-of-Dentistry
- 23 Preparing-reports-for-publication-and-responding-to-reviewers-comments\_2006\_Journal-of-Clinical-Epidemiology
- 24 Strategies-for-Getting-Published-in-High-Impact-Dermatology-Journals\_2008\_Actas-Dermo-Sifiliográficas (English Edition)
- 25 The-triad-of-scientific-publication-Reading-writing-and-reviewing\_2013\_Forensic-Science-International-Genetics-Supplement-Series Dermo-Sifiliográficas-English-Edition-
- 26 Writing-and-publishing-a-scientific-paper-Facts-Myths-and-Realities\_2015\_Medical-Journal-Armed-Forces-India
- 27 Writing-and-Publishing-Scientific-papers\_2012\_Gastroenterology



## Appendix II: Use of “English” in the Advice corpus

A:	Used as an adjective (12 occurrences)		
T3	Responding to reviewers series does not address	English	spelling and grammar. General suggestions
T11	scientists whose native language is not	English	. Participants learn that the burden of
T13	imprisoned in the wretched conventions of	English	society, without genius, wit, or knowledge
T13	original poetic language in the most robust	English	tradition but who seems to have been arrested
T14	articles. All academic institutions should have	English	departments so it is best for manuscripts
T14	thoroughly checked by persons with good	English	skills before being submitted to a journal
T15	journal, so it follows that it uses British	English	spelling and phraseology. If you are not
T15	example, “in German” in brackets after the	English	translation. General comments Several points
T16	your paper should be checked by a native	English	speaking scientist for grammar, clarity
T16	Lynne Truss. 13 If you do not have a native	English	speaker who can review your paper, read
T24	authors often give too much priority to	English	language journals, overestimate the actual
T24	Priority given to the language of publication (	English	) 1. Williams HC. How to reply to referees

**B. Used as a noun (21 occurrences)**

T1	many of whom are not native speakers of	English	.	With the impact of the Internet, even
T1	the message. Put the Action in the Verb In	English	,	verbs are where the action is. If the
T11	or abroad, are nonnative speakers of	English	[11, 12], or	lack scientific writing experience
T11	granting agency and that idiomatic American	English	is used	correctly. Finally, the manuscript
T11	helpful for those who are not fluent in	English	. Fourth,	concise information on the ethical
T13	Fowler, H. W. A Dictionary of Modern	English	Usage.	2nd ed., revised by Ernest Gowers
T13	M. The Careful Writer: A Modern Guide to	English	Usage.	New York, New York: Atheneum, 1965
T13	Taboos, Bugbears and Outmoded Rules of	English	Usage.	New York, New York: Simon &
T13	Theodore M. Doss, Don'ts & Maybes of	English	Usage.	New York, New York: Times Books,
T13	The Heritage Illustrated Dictionary of the	English	Language,	Volumes I and II. Boston,
T14	the original source of the idea(s).	English	is the	international language of science
T14	to disseminate work throughout the world,	English	is the	language in which the manuscript
T14	which the manuscript must be written. Poor	English	skills	are no excuse for plagiarising slabs
T14	scientific methodology Clear aim(s) Simple	English	–	correct grammar and succinct style Clear
T15	If you are not certain what British	English	is,	a good start is to make sure that your
T15	make sure that your version of Word uses “	English	(U.K.)	” or “English (United Kingdom)” as
T15	version of Word uses “English (U.K.)” or “	English	(United	Kingdom)” as its default setting
T15	papers published in languages other than	English	should	be translated and followed by the
T16	which I have slightly modified (Table 2). If	English	is	not your first language, your paper should
T24	● of these, 12 (23.5%) are published in	English	. Spanish	journals are of great value, not
T24	This stubborn insistence on publishing in	English	is	misplaced, as there are probably many



## Appendix III: Word Sketches of “English”

### A. Used as an adjective (12)

<u>adj_subject</u>		
	<u>1</u>	8.80
language	<u>1</u>	12.68

<u>modifier</u>		
	<u>1</u>	0.30
not	<u>1</u>	5.41

<u>modifies</u>		
	<u>10</u>	2.90
spelling	<u>2</u>	11.99
translation	<u>1</u>	11.41
tradition	<u>1</u>	11.09
society	<u>1</u>	10.75
speaker	<u>1</u>	10.75
scientist	<u>1</u>	10.35
department	<u>1</u>	9.95
skill	<u>1</u>	9.67
journal	<u>1</u>	6.11

<u>and/or</u>		
	<u>5</u>	3.10
native	<u>2</u>	13.19
robust	<u>1</u>	12.41
British	<u>1</u>	12.19
good	<u>1</u>	11.00

<u>adj_comp_of</u>		
	<u>1</u>	4.40
be	<u>1</u>	5.10

### B. Used as a noun (21)

<u>object_of</u>		
	<u>2</u>	0.80
misplace	<u>1</u>	12.68
use	<u>1</u>	7.36

<u>subject_of</u>		
	<u>6</u>	4.30
be	<u>6</u>	6.96

<u>modifier</u>		
	<u>6</u>	0.90
Idiomatic	<u>1</u>	12.19
Modern	<u>1</u>	11.99
Simple	<u>1</u>	11.99
world	<u>1</u>	11.83
British	<u>1</u>	11.00
American	<u>1</u>	9.57

<u>modifies</u>		
	<u>3</u>	0.50
Language	<u>1</u>	13.00
verb	<u>1</u>	10.91
skill	<u>1</u>	9.95

<u>and/or</u>		
	<u>2</u>	0.70
world	<u>1</u>	12.41
verb	<u>1</u>	12.19

<u>pp_obj_of</u>		
	<u>5</u>	5.30
speaker	<u>2</u>	13.19
Maybes	<u>1</u>	12.41
Rules	<u>1</u>	12.41
Dictionary	<u>1</u>	12.19

<u>pp_obj_in</u>		
	<u>4</u>	9.90
fluent	<u>1</u>	12.68
Verb	<u>1</u>	12.68
publishing	<u>1</u>	11.54
publish	<u>1</u>	9.24

<u>predicate_of</u>		
	<u>3</u>	31.80
language	<u>3</u>	13.99

<u>pp_obj_to</u>		
	<u>1</u>	3.80
Guide	<u>1</u>	13.41

<u>pp_obj_than</u>		
	<u>1</u>	37.70
other	<u>1</u>	12.68

# CHAPTER TWO

## LEXICAL DENSITY IN SCIENTIFIC RESEARCH GENRES

ELIZABETH ROWLEY-JOLIVET

### 1. Introduction

A key concept in SFL is that of language behaviour as choice: when we use language, we make meaning by choosing not only what sort of meaning we wish to express (or what functions we want to perform), but also what sort of wording, or lexicogrammatical expression, we encode the meaning in. These choices are to a large extent constrained or influenced by contextual factors that make one set of meanings more likely or appropriate to be expressed than another—factors such as communicative purpose and genre, the role relationships between the participants, and the mode of expression, speech *vs* writing. Context and language are in a reflexive relationship, and mutually construe each other, with the language simultaneously reflecting and constructing the type of context or social system.

The context, concepts and methods of science have changed radically since the moment when English began to be used regularly as a language of scientific communication in the late seventeenth century, and scientific discourse has developed in response to these changes, as is clearly shown by Banks (2008). Addressing this evolution from a diachronic perspective, he shows not only that the written language has changed over time but that it changes as a function of the context of scientific research in which it is produced in order to meet the communicative needs of its practitioners. The result, at the present time, is a language—in writing at least—characterized lexically by special, complex vocabularies or technical taxonomies, and grammatically by certain constructions that are favoured over others, such as a preference for nominalization to construct these taxonomies, extended pre- and post-modified noun phrases, declarative sentences, and constructions

enabling an ‘objective’ stance such as the passive voice and other impersonal structures (Halliday and Martin, 1993). The co-occurrence of these features results in a high lexical density, usually entailing a heavy information load as lexical items are content-carrying words. While it is recognized that writing in general tends to be lexically denser than speech (Halliday 1985), this is particularly marked in scientific language where the specialized lexicogrammar enables scientists to construct an alternative interpretation of the physical world to that provided by the commonsense language of everyday spontaneous speech.

Scientists, and academics in general, do not only write research articles and monographs, however. They also produce many other genres, both spoken and written, for the various communicative needs of their activity. We are much less well informed, however, about certain features of these other genres, for which the characteristics of the research article (RA) are likely to be a poor yardstick. In the present study, I will focus on one specific feature, lexical density (LD), across a range of scientific genres in English, in an attempt to fill in the picture and to provide some reference figures of LD that can then be used for research, training, or evaluative purposes. While there are a few figures for LD scattered throughout the literature, no systematic comparative study of the main research genres appears to have been carried out.

Information about LD would be useful, however, for training purposes in academic literacy, whether in secondary or higher education. Several studies, many of them by SFL scholars, have shown for example that coping with the language typically used in school textbooks is challenging for students (Christie and Derewianka 2008) due to the “high lexical density, abstraction, technicality, expanded noun groups, embedded clauses, and hierarchical structure” (Fang et al. 2006, 249) of these texts. Similarly, university students also encounter problems mastering the written academic register (Snow and Uccelli 2009) due to its very high information load and dense packing of information. Additional features, particularly in science, are the high degree of technicality of the lexicon and the authoritative, distanced authorial stance expected in writing, requiring knowledge of the appropriate lexico-grammar to construe scientific knowledge (Fang 2005):

Ideationally, the nominalizing grammar [of science] creates a universe of things, bounded, stable and determinate; and (in place of processes) of relations between things. Interpersonally, it sets itself apart as a discourse of the expert, readily becoming a language of power and technocratic control. In both aspects, it creates maximum distance between technical scientific knowledge and the experience of daily life. (Halliday 1998, 228)

Poor academic literacy will limit the student's access to, and subsequent progress in, higher education. There is therefore great interest among teachers, and particularly ESP trainers, for methods that enable students' academic literacy to be measured in order to chart their progress or weaknesses, and to help them handle the reading, writing and listening demands of their education. Lexical density is one such measure, and has been applied in various studies.

Following a survey among international students at British universities which found that their greatest area of concern was taking lecture notes at speed, Nesi (2005) studied the lexical density of university lectures from the BASE corpus and compared it with EAP training materials in listening comprehension. She found evidence in the authentic lectures of a link between speed of delivery and LD: lecturers produce faster and sparser text when they do not expect students to take notes (e.g. passages with predictable content, anecdotes). The recorded extracts used in conjunction with published EAP teaching material, however, did not reflect this variation of purpose and style of authentic lectures, and differed particularly in terms of density from those parts of authentic lectures that demand detailed note-taking, providing poor training for the real-life situation. In a corpus-based study of Chinese learners, Wu and Zou (2009) used lexical density to assess whether students were aware of the register differences between spoken and written English, and came to the negative conclusion that compared with data from native speakers, these EFL learners' speech and writing in English did not show strong register features. Yanmei (2003), in another study of Chinese university students, found on a more positive note that their lexical density in writing grew with the development of discourse competence and concluded that LD is a valid and reliable means for the assessment of discourse competence under controlled conditions.

Lexical density has also been used to measure the comprehensibility of documents targeting the general public and can therefore be a useful tool to assess the appropriacy for the receivers. Gemoets et al. (2004), for example, assessed the readability of consumer health information, using LD as one of their measures, and found that the best readability scores were obtained by texts with the lowest LD. Miller et al. (2007), in another study of health-related texts, concluded that consumer health web pages are not using appropriate language for their target audience, but pointed out that one also needs to take into account the technicality of the lexis used, not only its density: the comprehensibility will vary depending on the level of specialized knowledge of the reader. In natural language processing (NLP), LD is also used to automatically identify and model topic development and

the rhetorical structure of a text by detecting significant changes in lexical density (Cheung et al. 2001).

Several factors have been found to affect lexical density. The medium of delivery (speech or writing) is a major factor, as already mentioned, but LD may vary in speech depending on the presence or absence of feedback (Ure 1971; Nesi 2005), on whether it has been prepared/rehearsed in advance or is spontaneous, whether it is monologue or dialogue (Stubbs 1986), or how accustomed the speakers are to public speaking (Stubbs 1986). In writing, among other factors, the genre has a strong influence on LD, as shown for example by Biber et al.'s (1999) findings for news, fiction, and academic writing. In the present study, I have excluded speech genres containing dialogue or feedback, but have attempted to cover a fairly wide range of both spoken and written academic genres.

The article is structured as follows. After a description of the corpus used for the study in Section 2, the following section looks at lexical density in order to provide reference figures of LD for the seven genres examined and highlight their major differences. To complement the overall LD analysis, two other measurable characteristics focusing on the verbal group were also analyzed: verbal density in Section 4 and subject pronoun use in Section 5, revealing some interesting correlations between the three aspects. A brief discussion of the findings concludes the paper.

## 2. Corpus and Method

Seven genres, two spoken and five written, were selected for this study (see Table 1 for details). As the aim was to provide reference figures that can enlighten us on the characteristic features of research genres, and that can be used by ESP trainers of students/novice researchers, all the genres were produced by established academics and are related to research activity; didactic genres (such as textbooks or student lectures, for example) were excluded. They can therefore all be considered to belong to the same *genre network*, defined by Swales (2004, 22) as “the totality of genres available for a particular sector (such as the research world) as seen from any chosen synchronic moment”. The lexical density analysis presented here reflects the situation in the early twenty-first century. Two corpus tools were used for the analysis: the text annotator UAM CorpusTool developed by Mick O'Donnell for semi-automatic part-of-speech tagging,<sup>1</sup> and the concordancer AntConc developed by Laurence Anthony.<sup>2</sup> All the query

---

<sup>1</sup> <http://www.corpustool.com/>

<sup>2</sup> <http://www.laurenceanthony.net/software/antconc/>



results and concordances were manually post-edited to discard irrelevant occurrences.

All the texts in the corpus were delivered or written by native English speakers, as far as could be ascertained, in various hard science disciplines and medicine. The two spoken genres, conference presentations (henceforth CPs) and Three Minute Thesis talks (3MTs), were transcribed from the video recordings, as were the conference slides. While not a genre in themselves, the slides are included here as a separate category since they represent the extreme pole of the various parameters investigated, as will become clear below. The remaining four written genres were downloaded from the relevant websites. The word count for the CPs includes the speaker's monologue only, not the following question time; in 3MTs, there is no question time. The word count for the research articles is after the removal of Abstracts, Table and Figure glosses and captions, acknowledgements, references, and appendices.

Table 1. The corpus

	Genre	Number	Word count	Average Length (w.)
<b>Spoken</b>				
1.	Conference Talks	20	48,314	2416
2.	Three-Minute Thesis Talks	15	7,464	497
<b>Written</b>				
3.	Research Articles	20	117,906	5895
4.	Abstracts	30	10,802	360
	- <i>PhD Abstracts</i>	(15)	(7,169)	(478)
	- <i>RA Abstracts</i>	(15)	(3,633)	(242)
5.	Journal Editorials	40	29,447	736
6.	Laboratory Protocols	20	8,639	432
7.	Conference Slides	257 text slides	7,965	31
<b>TOTAL</b>			<b>230,537</b>	

Among the 20 CPs, 11 come from the 10<sup>th</sup> Novatech conference, a triennial international conference focusing on hydrology and the environment, 7 from an international conference on plasma chemistry and the remaining 2 from a symposium on oceanography. The scientific

disciplines covered by the 3MTs and the Abstracts of the corresponding PhD theses range from biomedicine to mechanics. The 20 research articles (RAs) and their Abstracts come from high-ranking journals in physics, chemistry and soil science. The Editorials sub-corpus was compiled from the medical journal *The Lancet*, and the sub-corpus of laboratory protocols from the sites of British and American research laboratories in biological engineering. The conference slides were extracted from research talks given at conferences in medicine, physics and geology; slides using the mathematical and visual semiotics were excluded, keeping only those slides that contained text.

The audiences targeted by some of the genres differ slightly. While the CPs, Slides, Abstracts, RAs and Protocols all target researchers in the specialized disciplinary field in question, the Editorials are addressed to a wider audience comprising three distinct readerships: medical researchers, socio-political actors whose decisions have an impact on medicine (governments, pharmaceutical companies, international health organizations, etc.), and medical practitioners who wish to keep up to date with progress in the field (Carter-Thomas and Rowley-Jolivet 2014). 3MT presentations are a new, competitive academic speech genre that gives PhD students the opportunity to present their ongoing research, in three minutes, to an academic but mixed disciplinary audience comprising doctoral students in other specialities and a panel of judges. To ensure comprehensibility and interest for the audience attending the contest, candidates are advised to “avoid jargon and academic language” and “convey your excitement and enthusiasm for your subject”.<sup>3</sup> The award-winning talks are then uploaded to the internet, making them available to the vast web audience (Rowley-Jolivet and Carter-Thomas forthcoming).

To calculate lexical density (henceforth LD), two categories of words are distinguished: lexical words (content words, or open-class items) and grammatical words (function words, or closed-class items). Content carrying words include nouns, the main part of the verb, adjectives, and some adverbs; non-content carrying words include prepositions, conjunctions, auxiliary verbs, some adverbs, determiners, and pronouns (Eggs 1994). The precise boundary between the two is slightly fuzzy, and as pointed out by Stubbs (1986), an automatic analysis, though generally reliable, efficient and capable of analyzing large datasets, will be unable to detect certain features (e.g. phrasal verbs will be analyzed as one lexical word (the verb), plus one function word (the postposition), rather than as a single lexical item; verbs such as *be*, *have* and *do* can be either auxiliary or full verbs depending on the grammatical contexts in which they are used). There is therefore a slight

---

<sup>3</sup> <https://threeminutethesis.uq.edu.au>

margin of uncertainty in the figures for LD reported in the literature. In an attempt to mitigate this uncertainty, two different methods were used for the calculation of LD in the present study: a home-made program developed some years ago, and the output of the UAM CorpusTool. Comparison of the results showed that agreement between the two sets of figures was very high (less than 1% difference, whatever the genre considered). Consequently, the LD figures reported here are those of the UAM CorpusTool, to allow comparison with other studies.

LD can be calculated in two ways. Halliday (1985) proposed a clausal method, in which LD is expressed as the number of lexical items per non-embedded clause: a text comprising 3 such clauses and 9 lexical items, for example, would have a mean LD of 3.0. While this method may be feasible for the very short extracts he selected to illustrate the principle, it is difficult to apply to large corpora or to extended passages of speech, where the degree of embeddedness heightens the analytical complexity. It is therefore the second method that is generally used, and that was adopted here, in which LD is calculated as the ratio of lexical words to the total number of running words in the text and expressed as a percentage (Eggs 1994, 60–61): an LD of 40% means therefore that forty percent of the running words are lexical items. LD is an indicator of the information density of a text, but cannot be entirely equated with it: not only is the degree of repetition of lexical items not taken into account, but the same text may represent a different information processing load for different readers.

### 3. Lexical Density

Table 2 and Figure 1 show the lexical densities of the 7 genres studied, presented in order of increasing LD.

Table 2. Lexical density in scientific research genres

Genre	Lexical Density (%)
Conference Talks	48.7
3MT Talks	49.7
Journal Editorials	57
Research Articles	60.4
Abstracts	60.4
- <i>RA Abstracts</i>	(63.3)
- <i>PhD Abstracts</i>	(58)
Protocols	67.6
Conference Slides	75



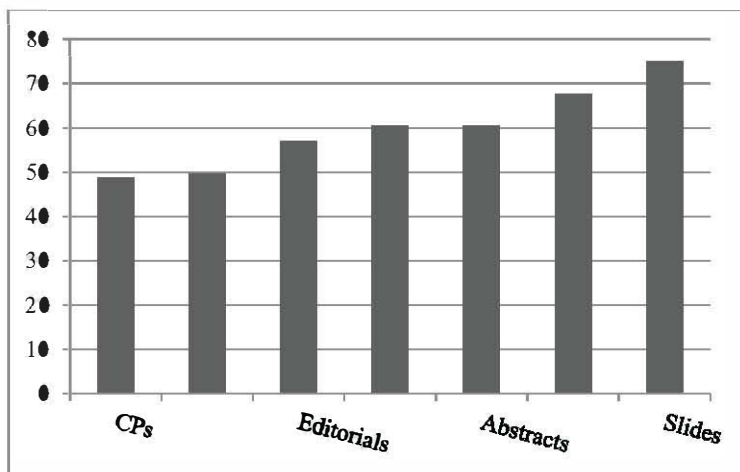


Fig. 1. Lexical density in scientific research genres.

Unsurprisingly, the two spoken genres (Conference talks and 3MTs) have a much lower LD than the written academic genres. The LD is nevertheless far higher than the language-in-action texts analyzed by Ure (1971), which averaged around 33%, showing that these prepared talks, delivered to an audience, are lexically much denser than what would probably be found in scientific language-in-action, for instance informal lab talk. The figure for CPs (48.7%) is practically identical to the figure of 48.4% found by Rowley-Jolivet (1998) in a CP corpus covering scientific disciplines different from those studied here, and very similar to the 46% reported by Hamilton and Carter-Thomas (2017) in their study of 15 scientific CPs by native English speakers. The convergence of these results seems to indicate a strong correlation between LD and genre: in other words, given the communicative purpose, type of audience, semiotic modalities used, and the cognitive constraints for both speaker and audience of processing research results in real time in the spoken mode, there appears to be both a lower limit (that one can situate around an LD of 45%) and an upper limit or cognitive ceiling (<50%) between which CPs delivered by English speakers will fall. Interestingly, Hamilton and Carter-Thomas (2017) found a slightly lower LD in scientific conference talks in French (42%), so the precise boundaries of LD in the CP genre may be language-specific to some extent, but further work is required to verify this.

The lexical densities of 3MTs and CPs (respectively 49.7 and 48.7) are highly comparable. There may be two opposing factors at work here,

however. In 3MTs, a factor that works towards increasing the LD is the total absence of the disfluencies—false starts, fillers, involuntary repetitions, etc.—of spontaneous speech. Due to the drastic time constraint, 3MT speakers have clearly prepared their presentations down to the very last word, as every second counts, leaving them no time for improvisation, back-tracking, etc. In some conference talks, however, disfluencies of various kinds are quite common, as illustrated in examples (1) and (2):

- (1) so first of all I'll go through this in I'll take about a minute but let me read this proposition well you can read it proposition that I've come up with and that we have come up with that we need to revise models that we have to incorporate microbial diversity and ecosystem function (CP)
- (2) okay we then combine those you can see here we've collected everything here's our we then I'm running out of time too quickly ah we then turn to the urbanCycle you can see here sorry urbanNet you can see here all these how do I go backwards? all these uh inputs into urban urbanNet and each of these this allows us to say where the water's coming from (CP)

While 3MTs are similar in some ways to recitation, in conference presentations some 'fresh talk' (Goffman 1981) does occur, indicating that the precise wording has been less fully prepared. Conversely, however, a factor that works towards decreasing the LD in 3MTs is the multidisciplinary audience: the very economical terminology of the speaker's specialized field cannot be used in this context, leading to reformulations and expansions that are lexically less dense (italicized in example 3):

- (3) So, the industry is also looking at *something called engine downsizing*. Now this is brilliant. *This is about getting the power of a big engine but from a much smaller engine.* (3MT)

In the conference talks, in contrast, the speaker addresses a peer audience of specialists and can use the terminology of the field without any further explanation or gloss. These two contrasting factors may cancel each other out, resulting in the almost identical LD observed between 3MTs and CPs.

There is a substantial gap between the LD of the two spoken genres and the first of the written genres, Editorials. With an LD of 57%, the Editorials approach the high LD of RAs and can be considered a sustained written register. The editorials appear to differ somewhat, however, from the news texts analyzed by Biber et al. (1999), who found that news articles had the highest LD (63%) of their four text-types, above that of RAs, and attributed this very high figure for newspaper articles to their communicative purpose:

“The main purpose of news reportage is to convey information, and preferably as concisely as possible; it is thus no surprise that it has the highest score for lexical density” (Biber et al. 1999, 62). The aim of Editorials is not primarily to convey information, however, but to comment on it, and this comment function not only contributes to the high LD overall of editorials (see Ex. 4) but may also explain the difference in LD between factual reportage and the more interpretative or analytical editorials (see Ex. 5). As pointed out by Francis (1989) in her comparative analysis of the lexical density of Themes in editorials and news reports, nominalization is a synoptic interpretation of reality: it freezes the processes and makes them static so that they can be talked about and evaluated. It is therefore well-suited to the rhetorical purpose of editorials. And indeed, one finds that heavy noun themes are frequently used in editorials to sum up known (or assumed known) information, leaving the writer space to then comment on it, as in example (4):

- (4) *One of the conclusions by the authors of the controversial research papers that showed how H5N1 avian influenza strains can be altered to pass between mammals* is perhaps unsurprising, but certainly gives pause for thought: the need to discuss the importance of the research earlier in the process of doing it. (ED)

On the other hand, another syntactic strategy that is also highly useful for evaluation, but that has the opposite effect in terms of the lexical density of the Theme, is extraposition (see Section 5 below on subject pronouns). Editorialists manipulate information structure to indicate their stance, downplay or give salience to certain points, often resorting to extraposition for this purpose:

- (5) *It is unimaginable that* such cuts will not affect service provision for the most vulnerable young people (ED)

The LD of the research articles (60.4%) is almost identical to that found by Rowley-Jolivet (1998) in Proceedings Articles in physics (59.3%) and by Hamilton and Carter-Thomas (2017) in the EILDA corpus (61%). As with the conference presentations, therefore, this homogeneity indicates a high degree of stability in the LD of this research genre in English, across a wide range of scientific disciplines. While there is a close intertextual relationship between RAs and CPs, in that many RAs and CPs deal with the same scientific content or piece of research, the change in mode, from written to spoken, clearly results in a radical reformulation of how the content is conveyed. The linguistic phenomena highlighted and analyzed at length by Halliday (1985, 1993, 1998), namely nominalization, expansion of the noun

group, and low grammatical intricacy are major contributors to the high LD of the RAs, as shown by examples (6) to (8). More recently, the corpus-based study by Biber and Gray (2010) confirmed that non-clausal modifiers embedded in lengthy noun phrases, and an absence of structural elaboration (e.g., paucity of finite subordinate clauses) are characteristic features of academic writing.

(6) nominalization of a material process:

(6a) *silanes are decomposing and producing silanol and silanolic acid* (CP)

(6b) *Silane decomposition occurs either by radicals or by electrons, leaving Si and (SiH<sub>2</sub>)<sub>2</sub>* (RA)

(7) expansion of the noun group, either by pre- or post-modification of the head noun. The head noun 'deposition' is both premodified by a heavy noun group (*high density plasma chemical vapor*) and postmodified by two prepositional phrases (*of SiO<sub>2</sub> films/ from silane-oxygen-argon mixtures*):

A detailed chemical kinetics mechanism was developed to predict clustering during high density plasma chemical vapor deposition of SiO<sub>2</sub> films from silane-oxygen-argon mixtures. (RA)

(8) low grammatical intricacy: in example (8a) from an RA, the extraposed clause contains a single declarative verb (*happens*) whereas in the corresponding passage from the conference talk (8b), the *that*-clause contains one deeply embedded relative clause (*we're simulating*), a syntactic break or disfluency (*we just assume that / we just ignore*) and a second relative clause (*that could be happening*).

(8a) it is assumed that nothing happens on the surface during this interval (RA)

(8b) in essence what we is then just assume that the typically millisecond or more between ion impacts for the layers of the size we're simulating we just ignore all the things that could be happening during this long time between ion impacts (CP)

Moving on to the next genre in Figure 1, the Abstracts, it can be seen that overall, the LD is identical to that of the RAs (60.4%). When broken down into the two subsets, however, of RA Abstracts and PhD Abstracts (see Table 2), a slightly different picture emerges: the LD of the RA abstracts, at 63.3%, is even higher than that of the RAs, whereas that of the

PhD abstracts is 5% lower, at 58%. It was to be expected that the RA abstracts would have a higher LD than the corresponding articles, as the abstract summarizes often extended passages of the RA in a single sentence, leading to the elimination of many details and of recurrent grammatical elements. The difference in LD between the RA and PhD abstracts is more intriguing, and together with the other differences noted below concerning verbal density and pronoun use, suggests either that the doctoral students have not yet fully mastered the art of writing scientific abstracts, or that PhD abstracts should perhaps be considered a specific genre, different from the RA abstract. As argued by Bordet (2011), PhD abstracts represent a specific communicative situation: the PhD applicant is not yet a member of the scientific community but is applying for membership; in this bid to gain access to the community, the role of the abstract is to act as a self-promotional text that demonstrates the candidate's personal competence and the legitimacy of his/her application. This specific purpose may have various linguistic consequences. For instance, there is a heavy use of linguistic items that highlight the novice researcher's accomplishments, his/her direct intervention in the research process, and the novelty of the thesis. Among these items, Bordet (2011) focused on the deictic *this*; its salience in PhD abstracts is confirmed by the data studied here, where *this* is the 10<sup>th</sup> most frequent word type, and accounts for 1% of all the tokens. We shall see below that other deictics are also used very frequently, and that this may well be part of the explanation for the lower LD in the PhD abstracts.

The remaining two genres, protocols and slides, have the highest LDs, at 67.6 and 75% respectively. Laboratory protocols, analyzed by Rowley-Jolivet (2012b) and Coutherut (2016), are a procedural or instructional genre, in which it is essential to ensure clarity of the successive steps in the procedure. Consequently, rather than using connectors, subordinators or relative clauses—all of which require function words—the instructions are presented as numbered lists, expressed by imperatives, which account for 74% of the verb forms in the protocols (9):

(9) Text layout in the Protocols.

3. Mix gently by inversion or finger tapping.
4. Place tubes on ice for 20 minutes.
5. Heat shock cells by placing tubes at 42 °C for 90 seconds. (Protocol)

Concision is also an essential consideration in protocols. This is achieved primarily by the ellipsis of a wide range of function words: determiners, in particular *the* (10a), pronoun complements (10a), some prepositions (10b),

pronoun subjects and verbal auxiliaries (10c). This increases the lexical density of the genre appreciably.

(10a) Place  $\emptyset$  solution on  $\emptyset$  magnetic stir plate and allow  $\emptyset$  to cool.

(10b) Add three volumes  $\emptyset$  ethanol

(10c) If two gel loadings  $\emptyset$  desired, aliquot portions into microfuge tubes and place  $\emptyset$  at  $-20^{\circ}\text{C}$  until  $\emptyset$  used.

In the final subset, text slides, the lexical density reaches the extremely high figure of 75%, whereas the corresponding spoken passages in the CPs, in which the speaker comments on the textual slides in question, have an LD of 48.5% (see Rowley-Jolivet 2012a for further details). The extreme economy of the slides is achieved by three main means: the use of scientific and typographical symbols instead of natural language for concepts, relations, and entities (Ex. 11); a telegraphic style, in which the message is reduced to the key nouns (see Ex. 16 below); and the ellipsis of many categories of function words, as in the protocols (Ex. 12). The result is a highly nominalised, condensed text – that poses no problems of comprehension, however, for the specialised international audience.

(11). Use of symbols on text slides in scientific conference talks:

Field	Written (slide)	Type
Physics	- vary $I$ , $Q_{\text{gas}}$ , $L_{\text{arc}}$	Physics symbols
Medicine	IORT	Field-specific acronym
Physics	$t = 1/\text{acc. Rate}$	Mathematics symbol
Geology	10500-11000 B.P.??	Punctuation marks to indicate hypothesis

(12) ellipsis of function words on text slides:

later experiments use  $\emptyset$  planar magnetron  
trend surfaces from 1st-6th order  $\emptyset$  fitted to  $\emptyset$  data and isobase maps  $\emptyset$   
produced

#### 4. Verbal density

Verbal density is a measure of clause length. Figure 2 shows the verbal density of the different genres, expressed as 1 finite verb every  $n$  running words.



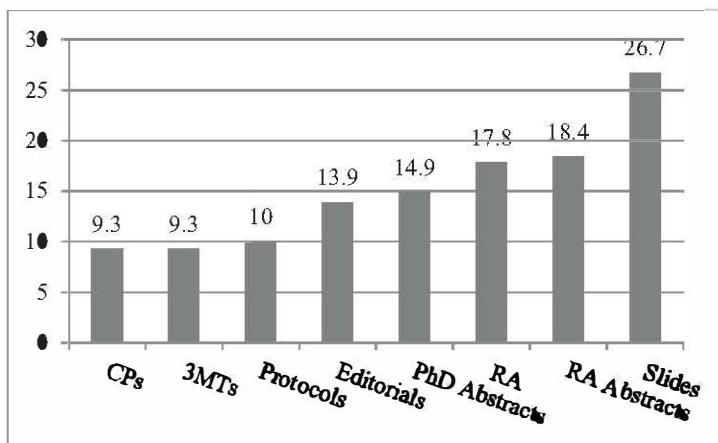


Figure 2. Verbal density in scientific research genres

RAs: 8 RAs sampled out of the total of 20; EDs: 20 editorials out of 40 sampled

Several features are worthy of comment. A first striking result is the identical verbal density of the two spoken genres, 3MTs and CPs, with relatively short clauses, averaging 9.3 words in length. This not only confirms the verbal density figure of 10 reported by Carter-Thomas (2004) for a small set of 9 CP Introductions, but also indicates that though 3MTs are a new genre, addressed to a rather different audience, they appear to obey the same constraints on clause length as the more established academic speech genre. Clauses in the RAs and RA abstracts, in contrast, are twice as long, averaging about 18 words, as non-finite verbal forms are heavily used (e.g. example 13b: *using, designed, incorporating*), whereas the spoken genre breaks the information into small chunks that can be easily processed by listeners in real time, as illustrated in (13a):

(13a) once we *start* marking the inflow with a tracer we *reveal* the mixing patterns within and then we *can seek to* gain insight into the nature of this system and how we *can design* it better (CP: 36 words, 4 finite verbs)

(13b) Baseflow (no rainfall) *was sampled* using a nested temporal scheme *designed to* adequately assess levels of variability in water quality (He et al. 2001) by incorporating four-time scales (hours, days, weeks and months) into the sampling strategy. (RA: 37 words, 1 finite verb)

Among the other written genres, the verbal density of the protocols is also very high, averaging 1 verb every 10 words. As in the spoken genres, the information is broken up into small bite-sized pieces, also for information processing reasons but of a rather different kind: the aim is to ensure that each step in the procedure is clearly marked out from the preceding and following ones so that no steps are skipped, merged, or duplicated. Each step is therefore given a separate line and its own verb in the description of the procedure (see (9) above).

As noted in section 3 on lexical density, there is again a difference between the RA abstracts and the PhD Abstracts, with a higher verbal density in the latter (1 verb every 14.9 words in the PhD abstracts versus every 18.4 words in the RA abstracts). Several factors may contribute to this. RA abstracts are highly compressed texts; as shown in Table 2, they are on average half the length of the PhD abstracts (242 words vs. 478). Moreover, as they are published in highly specialized journals that focus on a specific research field, authors can count on extensive shared knowledge with their readership and can thus often dispense with many general introductory remarks on the topic of the article or definitions of terms. As RA abstracts are also generally structured, there is no need to detail the different sections of the article. PhD abstracts, in contrast, are not only more extended, giving the author more space for introducing background information of a less technical nature, they also traditionally include a summary of the structure of the thesis. The impact of these features on verbal density can be clearly perceived in the following extracts from the PhD abstracts:

- (14a) [background] Human Immunodeficiency Virus, or HIV, *is* a virus that *attacks* and *destroys* the immune system, leaving you unable to defend yourself against infection. Though people infected with HIV *can live* many years without symptoms, the immune system *is* eventually so *damaged* that you *become* fatally ill, and *are considered* to have *Acquired Immunodeficiency Syndrome*, or AIDS.
- (14b) [definition] **Once a damper simulation was designed, a methodology for simulation parameterisation was required; parameterisation is the process of improving simulation performance through iterations of estimated parameters.**
- (14c) [structure of thesis] Chapter 1 *reviews* past and current approaches in NP fate modeling, [...] Chapter 2 *presents* a sediment diagenesis model that explores [...] Chapter 3 *presents* a coupled hydrologic, agricultural, and water quality model [...] Chapter 4 *compares* [...] Chapter 5 *summarizes* key findings



These can be compared with the first sentence from an RA abstract (Ex.15), which plunges the reader directly into the specific research topic and provides no gloss on the specialised terminology:

- (15) Two-dimensional direct numerical simulations *are performed* to investigate the non-linear dynamics of low Lewis number premixed flames, in the context of a two-step chain-branching chemistry model.  
(RA Abstract)

A final noteworthy feature is the extremely low verbal density of the conference slides, which are again at the extreme pole, with one verb every 26 words. This is mainly due to the extensive use of nominalisation of verbal processes: finite verbs are extremely rare on the slides, whereas the accompanying verbal commentary by the speaker uses mainly congruent forms, as illustrated in (16):

- (16)
- |                                     |  |
|-------------------------------------|--|
| <i>Text on slide</i>                | <i>Spoken commentary</i>   |
| Layer rotation for depth perception | What we'll do is take the layer, simply rotate it because it's difficult to see depth when you're just looking at the side layer |

## 5. Subject pronoun use

The choice of type of subject–noun group or pronoun–has a strong impact on lexical density, as pronouns are not only function words but also, apart from a few rare cases, cannot be pre- or post-modified by prepositional phrases, extended noun groups, etc. In a text with a large number of pronoun subjects, processes are expressed congruently; the text is therefore likely to have a lower lexical density than one in which most verbal subjects are noun groups. This is confirmed by Figure 3 which shows the percentage of finite verbs with a pronoun subject for each of the subsets, in increasing order. As can be seen, this figure is almost a mirror image of Figure 1 on lexical density: the two genres with the lowest lexical density (CPs and 3MTs) have the highest proportion of pronoun subjects, while the four subsets with the highest LD (Slides, Protocols, RA Abstracts and RAs) have the lowest proportion of pronouns as clausal subjects.

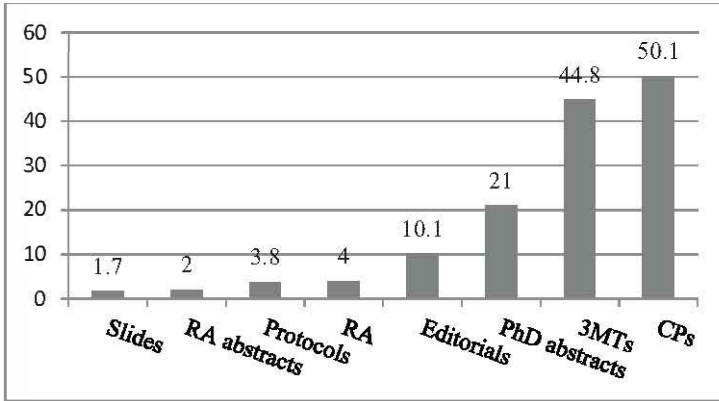


Figure 3. Percentage of finite verbs\* with a pronoun subject

\*RAs: 8 RAs sampled out of the total of 20; EDs: 20 editorials out of 40 sampled

† In the two spoken genres (3MTs & CPs), false starts and involuntary repetitions (e.g. I, I want...) were counted as a single occurrence

Table 3 gives the breakdown of the pronoun subjects, expressed as percentages of the total number of subject pronouns in each subset.

Table 3. Subject pronoun use across the 7 scientific genres

\*RAs: 8 RAs sampled out of the total of 20; EDs: 20 editorials out of 40 sampled

%	Slides	RA Abstracts	Protocols	RAs*	Editorials*	PhD Abstracts	3MTs	CPs
I	0	0	9	0	0	30	23.7	14.2
You	0	0	33.3	0	0	1	21.7	21.2
He /she	0	0	0	0	9.2	0	0.8	0.8
It	100	25	39.5	61.2	54.6	43	16	16.1
One	0	0	6	1	0	0	0	0.8
We	0	50	3	30.5	18.5	17	28.7	41
They	0	25	9	7.5	17.6	10	9	5.8
<i>Number of subject pronouns</i>	5	4	33	85	108	101	359	2608

These data raise several interesting points, enabling a finer-grained understanding of the factors that contribute to lexical density in the different

genres. On conference slides, many features of the interpersonal metafunction are extremely rare, or even absent, as it is the accompanying oral commentary that fulfils this function (Rowley-Jolivet 2012a). This is reflected here in subject pronoun use: the only subject pronoun used is the impersonal *it*, and even then, very sparsely (5 occurrences), given the scarcity of finite verbs on the slides (see section 4). Subject pronouns are also rare in the protocols, accounting for only 3.8% of clausal subjects, due to the overwhelming use of the imperative; protocols are however the only written genre in which *you* features as a subject pronoun, as protocol authors address their readers directly and therefore need the second person pronoun for certain comments or subordinate clauses (Ex. 17):

(17) When *you* are certain that the flame on the forceps has extinguished, repeat 2 times. (Protocol)

Unsurprisingly, a very low percentage of pronoun subjects is also found in the RAs (4%), mainly the impersonal *it*, the only personal pronoun used being *we*, generally exclusive *we* referring to the authors of the RA. The percentage of personal pronoun subjects may vary, however, depending on the discipline (Hyland 2001; Lafuente-Millán 2010). The Editorials are the only genre in which the third-person personal pronouns *he*, *she* and *they* are used to any significant extent, reflecting both the fact that medicine deals with human subjects (patients and physicians), and the focus mentioned above (see section 2) on the socio-political actors whose decisions have an impact on medicine; these actors regularly feature as the active agents of verbal processes in the Editorials:

(18) [Obama] *He* has also called for companies, universities, foundations, and philanthropists to join the project, which *he* hopes will generate thousands of jobs (ED)

Table 3 also shows that *it* accounts for over half the subject pronouns in editorials: in many cases, *it* does not refer to inanimate scientific entities (as in the protocols or RAs, for instance), but is the dummy subject of an extraposed clause:

(19) *It is crucial, however, that* America uses the momentum generated by the terrible events of the past few years to keep a conversation about gun control, and the prevention of gun violence, going. (ED)

As shown in Carter-Thomas and Rowley-Jolivet (2017), extraposition is a very useful informational structure in editorials: by placing the writer's point of view in the separate *it* clause, it enables the expression of very

critical comments and strong recommendations whilst also maintaining a polite distance, in an indirect display of power over the reader.

A striking finding is the huge difference in subject pronoun use between the RA Abstracts and the PhD Abstracts, which differ by a factor of 10 (respectively, 2% of finite verbs with pronoun subjects vs 21%). While the previous sections have revealed some differences in lexical and verbal density between the two subsets, here the difference is extremely marked, not only in terms of frequency but also in terms of the different pronouns favoured in each subset. The PhD abstracts are the only written genre with a significant proportion (30%) of the first-person pronoun *I*. This is in part due, of course, to the fact that PhD theses are single-authored texts; it also appears to indicate, however, a very strong authorial presence by the doctoral candidates, in line with the 'self-promotional' purpose of PhD abstracts proposed by Bordet (2011), and illustrated in (20):

(20) In this thesis *I explore* unsupervised methods for inducing parts of speech across languages. *I discuss* the challenges in evaluation of unsupervised learning and at the same time, by looking at the historical evolution of part-of-speech systems, *I make the case* that the compartmentalised, traditional pipeline approach of NLP is not ideal for the task. (PhD Abstract)

On closer inspection, however, it turns out that only 3 out of 15 PhD abstracts (20%) adopt this personal stance. In the remaining 12, the other syntactic options traditionally used in research writing for presenting the work accomplished – the passive, exraposition, or *we* to refer to the research team hosting the doctoral student – are used instead. This heterogeneity in authorial stance suggests that the discourse conventions for PhD abstracts are perhaps less rigid, or less well-established, than those of RA abstracts. A larger and more multidisciplinary corpus would be necessary to check this.

Lastly, yet again it is the speech/writing divide that leads to the greatest difference in subject pronoun use: in both the spoken genres (CPs and 3MTs), up to half of clauses have a pronoun subject, confirming the figure of 48% found in Rowley-Jolivet and Carter-Thomas (2005b) for CP introductions. The presence of a live audience leads speakers even in hard disciplines to take a discursive stance characterized by a higher degree of personal involvement and interaction with the receivers. Moreover, Table 3 shows almost identical percentages in both CPs and 3MTs for the different pronouns apart from *I/we*, for the obvious reason that the research presented in conference talks is generally a collaborative effort, unlike the more solitary doctoral research.

## Summary and Concluding Remarks

The three measurable characteristics examined in the previous sections—lexical density, verbal density, and subject pronoun use—have brought out some salient differences between the scientific genres studied here. While the LD of all these genres is noticeably higher than that of conversation, there is nonetheless a very wide range, from around 48% for the two spoken genres up to 75% in conference slides, with the other four written genres, Editorials, RAs, Abstracts and Protocols, exhibiting a progressively increasing LD between these two extremes. Grouping the LD, verbal density and subject pronoun figures together for each of the seven genres (see Figure 4), several general conclusions can be proposed.

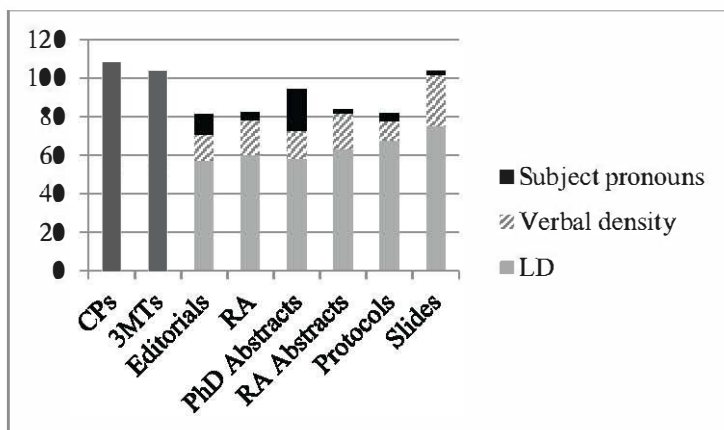


Figure 4. Aggregated figures for lexical density, verbal density and subject pronouns across the 7 scientific genres

The first striking fact is the very strong similarity between the two spoken genres, which have quasi-identical low lexical densities, high verbal densities and prominent use of subject pronouns, indicative of a congruent rather than a nominalized lexicogrammar. This is a little unexpected, as they differ in several respects: the new academic genre of 3MTs are extremely short (3 min vs 20 min CPs), and are delivered by novice, not confirmed, researchers, to a less specialized audience than that of the CPs. The spoken mode therefore appears to be the determining factor, overriding these differences. This is substantiated by the fact that when the figures obtained here for the CPs are compared with those available in the literature for other scientific disciplines, one finds a remarkable stability in the LD (around

48%). This highlights the interdependence of context and language and indicates that the term of language ‘choice’, as pointed out by Thompson, “does not necessarily imply a conscious process of selection by the speaker” (1996, 8): in fact, it appears to be rather the genre and mode that choose the lexical density on the speaker’s or writer’s behalf.

Turning to the written genres, this stability is confirmed for the RAs, as a very regular LD of ~60% is found in the data analyzed here and in other published figures, again despite disciplinary differences. The Editorials, in contrast, though published alongside the RAs in the same journal, differ in all three features: the LD is somewhat lower, while clauses are appreciably shorter (verbal density is higher), with many more congruent wordings (higher proportion of subject pronouns). An interesting finding that calls for further study is the differences observed between the PhD Abstracts and the RA Abstracts. PhD Abstracts have a lower LD, shorter average clause length, and much greater proportion of clauses with pronoun subjects, in particular the first person *I*. The remaining two genres, Protocols and text Slides, share a specific feature which boosts their LD up to very high figures (respectively 67.6 and 75%), namely a telegraphic style involving the ellipsis of a large number of function words. They also both make sparse use of pronoun subjects, though for different reasons: a heavy use of imperatives in the Protocols, and a high degree of nominalization on the Slides. As a result, their verbal densities are very different.

Needless to say, the three features examined here are far from exhausting an analysis of the factors contributing to lexical density and how information load is handled in different discursive contexts and genres. For reasons of space, I have not dealt with noun pre- or post-modification, for example, nor with dependency phenomena such as para- and hypotaxis. There are, however, other rhetorico-pragmatic factors that are perhaps less easily quantifiable but that also impact on informational density. These can be illustrated by comparing the following extracts from a conference talk on marine biology and the corresponding research article (Ex. 21), where there is a 20% difference in the lexical density of the spoken and written versions.

(21)

*CP*  
 so this is rather a cute story  
 so what my student found  
 was this that that diatoms if  
 you don't have bacteria silica  
 does not dissolve okay he  
 studied it to the point where  
 he could show that

*RA [citations removed]*  
 Living diatoms protect their  
 silica frustules from dissolution  
 by surrounding them with an  
 organic matrix. We previously  
 found that bacteria regulate Si  
 regeneration from  
 experimentally lysed diatom

colonizing bacteria express proteases which slough off the proteoglycan glycoprotein from the surface of the diatoms' silica shell that protects silica from dissolving once that happens silica dissolves rapidly no bacteria no dissolution so what this paper says now is that if you look at the temperature regulation of this biochemistry you find that you can explain in these terms that the accumulation of opal in Antarctic is explained by the temperature low temperature effect on the biochemistry of removal of organic coating and dissolution of silica okay

detritus and from natural diatom blooms [...] by colonizing and degrading the protective organic matrix surrounding diatom frustules. Bacterial mediation of dissolution suggests that the temperature probably controls not only the chemical depolymerization of silica but also the rate of the bacterial removal of the protective organic matrix.

In (21), a structure that increases the verbal density and lowers the LD in the conference talk is, for example, the use of *wh*-clefts (*what my student found was...*, *what this paper says is...*). In the data analyzed here, of the total 169 *wh*-clefts found in the whole corpus, 98% (166) occur in the two spoken genres, particularly in the CPs, where they are used as an information packaging strategy to highlight the division between Given and New, giving salience to the New, and as an interactive strategy (Rowley-Jolivet & Carter-Thomas 2005a). Another feature is repetition of the key ideas or claims that the speaker wishes to communicate: as the CPs are delivered, and processed by the audience, in real time, repetition is a necessary oratorical device to ensure that the main message gets home. In (21), for example, we find early on in the passage the congruent utterance *if you don't have bacteria silica does not dissolve*, which is reformulated later in nominal terms as *no bacteria no dissolution*. In addition to providing punchy summaries of preceding developments (*no bacteria, no dissolution*), speakers also signal in advance what is to come, preparing the listener for the type of information they are about to hear (*so this is rather a cute story*): in this case the research finding is presented in the form of a narrative with active human participants who are the agents of processes (*my student found, he studied, he could show*). These different lexicogrammatical and rhetorical choices construe two different contexts, academic speech and academic writing, in which the same research investigation is radically



recontextualized to take into account the discursive situation and the audience's needs and expectations. In this process of recontextualization, lexical density, as has hopefully been shown in the previous sections, plays an important role.

## References

- Banks, David. 2008. *The Development of scientific writing*. London: Equinox.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Pearson Education Ltd.
- Biber, Douglas, and Bethany Gray. 2010. "Challenging stereotypes about academic writing: Complexity, elaboration, explicitness." *Journal of English for Academic Purposes*, 9: 2-20.
- Bordet, Geneviève. 2011. *Étude contrastive de résumés de thèse dans une perspective d'analyse de genre*. Doctoral dissertation, Université Paris-Diderot-Paris VII.
- Carter-Thomas, Shirley. 2004. "Specialised syntax for specialised texts? A comparison of the preferred syntactic patterns in proceedings article and conference presentation introductions." *Actes du Colloque GLAT 2004*: 11-21.
- Carter-Thomas, Shirley, and Elizabeth Rowley-Jolivet. 2014. "A syntactic perspective on rhetorical purpose: the example of *if*-conditionals in medical discourse." *Iberica*, 28: 59-81.
- Carter-Thomas, Shirley, and Elizabeth Rowley-Jolivet. 2017. "Maintaining a dominant voice: a syntactic analysis of the way power is wielded in medical editorials." In: *Power, persuasion and manipulation in specialised genres: providing keys to the rhetoric of professional communities*, edited by María Angeles Orts, Ruth Breeze and Maurizio Gotti, 71-99. Bern: Peter Lang.
- Cheung, Lawrence Y. L., Tom B. Y. Lai, Benjamin K. Tsou, Francis C. Y. Chik, Robert W. P. Luk, and Oi Yee Kwong. 2001. "A preliminary study of lexical density for the development of XML-based discourse structure tagger." *Proceedings of the 1st NLP and XML Workshop*: 63-70.
- Christie, Frances and Beverly Derewianka. 2008. *School discourse: Learning to write across the years of schooling*. London: Continuum.
- Couterut, Margaux. 2016. *Les Textes procéduraux en anglais : création d'une échelle de prototypicalité*. PhD Dissertation, Université Paris Diderot.



- Eggs, Suzanne. 1994. *An Introduction to Systemic functional linguistics*. London: Pinter Publishers.
- Fang, Zihui. 2005. "Scientific literacy: A systemic functional linguistics perspective." *Science Education* 89, no. 2: 335–347. doi: 10.1002/sce.20050
- Fang, Zihui, Mary J. Schleppegrell, and Beverly E. Cox, 2006. "Understanding the language demands of schooling: Nouns in academic registers." *Journal of Literacy Research* 38, no 3: 247–273.
- Francis, Gill (1989) "Thematic selection and distribution in written discourse." *Word* 40, no.1-2: 201–221. doi: 10.1080/00437956.1989.11435804
- Gemoets, Darren, Graciela Roseblat, Tony Tse and Robert Logan. 2004. "Assessing readability of consumer health information: an exploratory study." In *Medinfo 2004*, edited by Marius Fieschi, Enrico Coiera and Yu-Chan Jack Li, 869–873. Amsterdam: IOS Press.
- Goffman, Erving. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.
- Halliday, Michael A.K. 1985. *Spoken and written language*. Oxford: Oxford University Press.
- Halliday, Michael A. K. 1993. "Some grammatical problems in scientific English." In *Writing science: Literacy and discursive power* edited by Michael A. K. Halliday and J. R. Martin, 69–85. London: The Falmer Press.
- Halliday, Michael A.K. 1998. "Things and relations: Regrammaticising experience as technical knowledge." In *Reading Science*, edited by J.R. Martin and Robert Veel, 185–235. London: Routledge.
- Halliday, Michael A.K. and J.R. Martin. 1993. *Writing Science: Literacy and discursive power*. London: The Falmer Press.
- Hamilton, Clive and Shirley Carter-Thomas. 2017. "Competing influences: the impact of mode and language on verb type and density in French and English scientific discourse." *Chimera. Romance Corpora and Linguistic Studies* 4, no. 1: 13–34.
- Hyland, Ken. 2001. "Humble servants of the discipline? Self-mention in research articles." *English for Specific Purposes* 20, no. 3: 207–226.
- Lafuente Millán, Enrique. 2010. "Extending this claim, we propose...! The writer's presence in research articles from different disciplines." *Ibérica*, no. 20: 36–56.
- Miller, Trudi, Gondy Leroy, Samir Chatterjee, Jie Fan, and Brian Thoms. 2007. "A classifier to evaluate language specificity of medical documents." *Proceedings of the 40th Hawaii International Conference on System Sciences*. IEEE, 134–140.

- Nesi, Hilary. 2005. "A Corpus-Based Analysis of Academic Lectures across Disciplines." In *Language Across Boundaries* edited by Janet Cotterill, and Anne Ife, 201–218. London: Continuum.
- Rowley-Jolivet, Elizabeth. 1998. *La communication scientifique orale: étude des caractéristiques linguistiques et discursives d'un genre*. PhD Dissertation, Université de Bordeaux 2.
- Rowley-Jolivet, Elizabeth. 2012a. "Oralising text slides in scientific conference presentations: A multimodal corpus analysis." In *Corpus-Informed research and learning in ESP: Issues and applications*, edited by Alex Boulton, Shirley Carter-Thomas, and Elizabeth Rowley-Jolivet, 137–165. Amsterdam: John Benjamins.
- Rowley-Jolivet, Elizabeth. 2012b. "Open science and the re-purposing of genre: An analysis of web-mediated laboratory protocols." In *Evolving genres in Web-mediated communication*, edited by Sandra Campagna, Giuliana Garzone, Cornelia Ilie, and Elizabeth Rowley-Jolivet, 127–150. Bern: Peter Lang.
- Rowley-Jolivet, Elizabeth, and Shirley Carter-Thomas. 2005a. "Genre awareness and rhetorical appropriacy: Manipulation of information structure by NS and NNS scientists in the international conference setting." *English for Specific Purposes* 24, 41–64.
- Rowley-Jolivet, Elizabeth, and Shirley Carter-Thomas. 2005b. "The rhetoric of conference presentation introductions: Context, argument and interaction." *International Journal of Applied Linguistics*, vol. 15, no 1: 45–70.
- Rowley-Jolivet, Elizabeth, and Shirley Carter-Thomas. Forthcoming. "Scholarly Soundbites: Audiovisual Innovations in Digital Science and their Implications for Genre Evolution." In *Science communication on the Internet. Old genres meet new genres* edited by María José Luzón and Carmen Pérez-Llantada. Amsterdam: John Benjamins.
- Snow, Catherine E., and Paola Uccelli. 2009. "The challenge of academic language." In *The Cambridge Handbook of Literacy*, edited by David R. Olson and Nancy Torrance, 112–133. Cambridge: Cambridge University Press.
- Stubbs, Michael. 1986. "Lexical density: A computational technique and some findings." In *Talking about Text*, edited by Malcolm Coulthard, 27–48. Birmingham: University of Birmingham Press.
- Swales, John M. 2004. *Research genres*. Cambridge: Cambridge University Press.
- Thompson, Geoff. 1996. *Introducing Functional Grammar*. London: Arnold.

- Ure, Jean. 1971. "Lexical density and register differentiation." In *Applications of linguistics* edited by George E. Perren and John L.M. Trim, 443–452. Cambridge: Cambridge University Press.
- Wu Jin, and Qing Zou. 2009. "A Corpus-based Register Study of Chinese Learners' Spoken and Written English: Lexical Density and Frequency." *Shandong Foreign Language Teaching Journal*, no.1.
- Yanmei, Gao. 2003. "Assessing Discourse Competence with LD (Lexical Density)." *Foreign Language Education* 5.

CHAPTER THREE

IMPREGNATION DE LA LANGUE  
PAR LA CULTURE :  
L'EXEMPLE DE L'ANGLAIS DE LA MEDECINE

PASCALINE FAURE

**Abstract**

The objective of this study is to demonstrate that English medical terminology is motivated on the one hand by the nature and constraints of Western medical practice (brevity and medical secrecy) and on the other hand by Anglo-Saxon culture (merit, science, travel, humour, food, and lifestyle). We propose therefore to correlate each constraint and each culture with a linguistic process (truncation and hypallage for the need to be concise, euphemism for the need to be discreet, anthroponymy for the culture of merit, orthonymy for the culture of scientificity, or toponymy for the culture of travel). Among these stylistic processes, metaphor plays an important role. More than simply a stylistic figure, it is a component of medical theory. Through the analysis of recent metaphors, we will show that current Western medicine is at an epistemological crossroads and seems to be moving, with SP medicine, towards highly technicised medicine, despite a resurgence of metaphors borrowed from ecology that could - or could have - given hope of a return to humanist medicine.

**Introduction**

La langue médicale anglaise présente un certain nombre de traits linguistiques caractéristiques qui sont déterminés par des nécessités propres à la pratique de la médecine occidentale tels que la concision, que l'on retrouve dans l'hypallage et la troncation, ou encore la discrétion, que traduisent les euphémismes, les codes et autres moyens de cryptage. Nous

analysons ces nécessités et leurs traductions dans la langue, dans notre première partie.

À côté de ces besoins spécifiques, les termes sont également façonnés par la culture dans laquelle la langue médicale s'inscrit. Cette culture définie comme étant « l'ensemble des traits distinctifs, spirituels, matériels, intellectuels et affectifs, qui caractérisent une société ou un groupe social<sup>1</sup> » évolue en fonction de paramètres politiques, économiques et scientifiques. Notre deuxième partie est consacrée à l'analyse de la culture – ou plutôt des cultures – dont la langue médicale anglaise se fait la traduction.

Les métaphores militaires sont parmi les plus anciennes de la langue médicale. Celle de l'économie de marché est plus récente. La métaphore de l'écologie est préconisée et commence à peine à (re)émerger. La guerre, l'économie de marché et l'écologie ainsi que leurs traductions dans la langue médicale anglaise sont traitées dans notre troisième partie.

Dans notre conclusion, face à l'émergence de nouveaux termes (*empowerment*, *eHealth*, *omics*, etc.) que nous avons relevés dans un certain nombre de publications médicales récentes, nous nous interrogeons sur l'apparition de nouvelles cultures médicales et plus particulièrement sur l'éventuel changement de paradigme que pourrait constituer la médecine 5P<sup>2</sup> et sa révolution annoncée.

Nous sommes partie de deux hypothèses selon lesquelles la terminologie médicale anglaise serait motivée par des contraintes propres à la nature et à la pratique de la médecine occidentale et qu'elle laisserait entrevoir les cultures professionnelles anglo-saxonnes au gré desquelles elle évoluerait. Les termes examinés dans cette étude sont issus de dictionnaires monolingues et bilingues anglais-français spécialisés, d'articles et de sites scientifiques médicaux, et d'ouvrages traitant de terminologie médicale en anglais, et ont été sélectionnés en fonction de la nature de leur motivation. L'objectif général de cette étude terminologique est de proposer une lecture de la terminologie médicale anglaise actuelle à travers le prisme des pratiques et des cultures professionnelles anglo-saxonnes.

## 1. La langue de la médecine, une langue contrainte

Comme toutes les langues de spécialité, la langue de la médecine s'est construite en fonction des besoins propres à la profession médicale. De la

---

<sup>1</sup> Définition proposée par l'UNESCO.

<sup>2</sup> La médecine 5P (en anglais, on parle plutôt de *P4 medicine*) est une médecine préventive, prédictive, participative, personnalisée, et pertinente. <https://www.telos-eu.com/fr/societe/revolution-numerique-et-medecine-5p.html>. Consulté le 09/01/2019.

nécessité de nommer aussi précisément que possible parties du corps, maladies, techniques et instruments pour éviter ambiguïté et contre-sens à l'obligation de s'exprimer vite et d'aller à l'essentiel afin de sauver les vies, la médecine s'est forgé des outils linguistiques qui devaient servir sa cause au gré des progrès scientifiques et des recommandations dans la pratique professionnelle. Même si Jean Hamburger (1982) lui reproche de peiner à suivre le rythme de la discipline qu'elle est censée accompagner, voire anticiper, la langue médicale a indubitablement évolué des théories hippocratiques dont on trouve encore des traces (*pituitary*, *phlegm*, *rheumatitis*, *lymph*, etc.) à la médecine génomique (*transcriptome*<sup>3</sup>, *mobilome*<sup>4</sup>, *proteome*<sup>5</sup>, etc.) basée sur l'utilisation de l'information génomique des individus pour le diagnostic ou le choix thérapeutique.

### 1.1. Troncation et concision

Lorsqu'il s'agit de sauver une vie humaine, en général, il n'y a pas de temps à perdre, et il est nécessaire d'être concis. Par conséquent, pour aller vite, la langue médicale anglaise use et abuse de raccourcis. Ces raccourcis peuvent prendre différentes formes parmi lesquelles la troncation est, sans nul doute, la plus courante.

La troncation peut se traduire par des aphérèses (*scope* à la place de *endoscope*), des syncopes (*appy* pour *appendectomy*) ou des apocopes (*gyn* pour *gyn(a)ecology*) (Faure 2012). L'haplogie ou hapaxépie, qui consiste à omettre des syllabes répétées ou proches, est un procédé de troncation également présent dans la langue médicale anglaise : *preventive* au lieu de *preventative*, *appendectomy* au lieu de *appendicectomy*, et *contraception* au lieu de *contraconception*. Il arrive parfois que l'ellipse concerne un lexème à part entière comme *cyto* présent dans *leukocytopenia* et qui a été perdu dans *leukopenia*.

On trouve aussi de nombreux amalgames : *spansule* (*spanning* + *capsule*), *urinalysis* (*urine* + *analysis*) et *fecalysis* (*fecal* + *analysis*), et des ellipses de mots entiers : *a physical* pour *a physical examination* et *a differential* pour *a differential diagnosis*.

Parmi les procédés abrégatifs, la siglaison est très certainement le plus couramment utilisé et probablement le plus décrié. En effet, l'anglais médical fourmille de sigles qui peuvent être des maladies (*AD* pour *Alzheimer disease*), des parties du corps (*ENT* pour *Ear Nose Throat* 'nez,

<sup>3</sup> Le transcriptome est l'ensemble des ARN issus de la transcription du génome.

<sup>4</sup> Le mobilome constitue les éléments mobiles du génome.

<sup>5</sup> Le protéome est l'ensemble des protéines exprimées dans une cellule.



oreille, gorge'), des examens (*FBS* pour *Fasting Blood Sugar* 'glycémie à jeun'), ou encore des services hospitaliers (*ICU* pour *Intensive Care Unit* 'unité de soins intensifs'). Ces sigles abondent sur les fiches d'examen clinique, qui doivent être remplies rapidement : *PC* pour *Present Complaint* 'motif de consultation', *OE* pour *On Examination* 'à l'examen', etc. On observe la présence d'un /x/ indicé à la suite de certains sigles. Elle indique qu'il s'agit d'un terme médical : *H<sub>x</sub>* pour *history* 'antécédents', *D<sub>x</sub>* pour *diagnosis*, *R<sub>x</sub>* pour *treatment*, *B<sub>x</sub>* pour *biopsy*, *F<sub>x</sub>* pour *fracture*, *Ab<sub>x</sub>* pour *antibiotics*, etc. Ce *x* proviendrait du symbole de Jupiter : ♃ (Dirckx 1983, 110).

Certains sigles peuvent s'inscrire dans le terme lui-même : *arbovirus* de *ARthropod-BORne virus* et *vipoma* de *Vaso-active Intestinal Peptide* + le suffixe *-oma*, une tumeur endocrine qui secrète du peptide vaso-actif intestinal. On peut même observer la transformation en un sigle de tout ou partie d'un énoncé : *HNPU* pour *Has Not Passed Urine* 'n'a pas uriné'. En dépit de leur commodité, on leur reproche leur caractère abscons, leur surabondance (Baue 2012) et leur ambiguïté. En effet, un même sigle renvoie parfois à plusieurs signifiés : *MI* pour *myocardial infarction* 'infarctus du myocarde' et *mitral insufficiency* 'insuffisance mitrale'.

Par souci de rapidité, la langue médicale anglaise recourt également à l'hypallage : *acute abdomen* ('abdomen aigu', c'est-à-dire 'un patient présentant un problème abdominal sérieux nécessitant une intervention rapide'), *liver panel* ('bilan hépatique', c'est-à-dire un dosage de certains éléments du sang afin de mettre en évidence un dysfonctionnement du foie), et *a perforated IUD* ('stérilet perforé', c'est-à-dire 'un stérilet qui a perforé l'utérus').

Par commodité, la langue médicale dérive. Cette dérivation peut s'effectuer tantôt sans modification morphologique [dérivation **0** ou 'conversion'] : le verbe *dose* 'donner un médicament' du nom *dose*, et le verbe *bag* 'ranimer à l'aide d'un ballon d'insufflation' du nom *bag* 'sac', tantôt avec une modification morphologique : *heparinize* 'mettre sous héparine' de *heparin* ou encore *therapeutize* 'traiter' de *therapy*. On dérive même les anthroponymes (l'adjectif *Swan-Ganz* signifie 'à l'aide d'un cathéter Swan-Ganz' comme dans *Swan-Ganz catheterization*, et *parkinsonian* de *Parkinson*) et les sigles (les verbes *DC* pour *discontinue*, *V-fib* de *Ventricular fibrillation* et *C-section* de *Caesarean section*) et les abréviations (le verbe *bronch* de *bronchoscopy*).



## 1.2. Cryptage et secret médical

Les sigles sont souvent utilisés comme moyen de cryptage lorsque le médecin ne veut pas que le patient comprenne : *EtOH* désigne un patient alcoolique, *Hi5* un patient séropositif (le 5 renvoie au V, le 5 en chiffre romain, de HIV) et *ToP* (de *Termination of Pregnancy*) une patiente qui se présente pour un avortement. Le sigle peut donner lieu à une dérivation : *a beemer* (de *BMI* ou *Body Mass Index*) pour désigner un patient obèse (Goldman 2014).

Afin d'éviter de blesser l'orgueil d'un patient, le médecin peut recourir à des euphémismes : *excessive soft tissue* est synonyme de 'obésité' ; *social injury of the rectum* désigne un patient qui se présente avec un objet dans le rectum ; *a drinking problem* est un autre terme pour *alcoholism* ; *a child with special needs* est un enfant handicapé mental ; *physically challenged* désigne une personne handicapée physique ; et *a space-occupying lesion* est une tumeur cancéreuse.

Dans un souci de ne pas inquiéter le patient, le médecin utilise parfois des barbarismes tels que *ditzel*, qui désigne une masse généralement sans gravité sur un cliché radiologique et qui viendrait de *ditsy* ou *ditzy* 'sans importance'<sup>6</sup> ; des termes empruntés à d'autres langues tels que le mot yiddish *schmutz* qui, en anglais médical, désigne une opacité anormale ; ou empruntés à la langue médicale technique tels que *heme* pour désigner le sang.

Parmi les codes utilisés dans les hôpitaux américains, on notera *404 moment*, qui désigne un patient dont on a égaré le dossier (de *404 Page Not Found*) ; *10th Floor Transfer* et *Room 13* indiquent que le patient est mort ; *Category 6 Patient*<sup>7</sup> désigne un patient qui n'a rien et qui fait perdre du temps à l'équipe soignante. En Grande-Bretagne, on déclenche le *code 99*<sup>8</sup> si on a besoin de ranimer un patient ou si l'hôpital va devoir gérer une catastrophe impliquant plus de 20 personnes. Aux États-Unis, on déclenche un *code blue* en cas d'arrêt cardio-respiratoire. Les codes peuvent donc varier d'un pays à l'autre. Ainsi, en Grande-Bretagne, si un patient est sous l'influence de stupéfiants, on déclenche un *code pink*. Aux États-Unis, ce même code signifie que le patient est atteint d'une maladie très contagieuse et, en France, qu'un enfant est en arrêt cardio-respiratoire.

<sup>6</sup> Source : Rogers, L. 2001. "Ditzels" Little Things Mean A Lot. Consulté le 09/01/2019. <http://www.ajronline.org/doi/full/10.2214/ajr.176.6.1761355>.

<sup>7</sup> Les catégories de triage vont de 1 (très grave) à 5 (bénin).

<sup>8</sup> Le numéro d'urgence est le 999 en Grande-Bretagne.

## 2. La langue de la médecine, une langue-culture

La nature de la médecine et les contraintes liées à sa pratique ne motivent pas à elles seules tous les termes médicaux. Une partie de ces termes est également motivée par les cultures professionnelles et a évolué à travers les siècles. En analysant la terminologie médicale, nous sommes en mesure de dégager les différentes cultures qui ont marqué la médecine occidentale jusqu'à nos jours.

### 2.1. Les anthroponymes et la culture du mérite

Parce que la médecine doit tout aux hommes qui l'ont inventée au fur et à mesure des découvertes qu'ils faisaient, souvent au mépris de leur vie, la langue de la médecine fait une place importante à l'anthroponymie et derrière elle, à la culture du mérite.

Il existe, dans la langue médicale anglaise, 6028 anthroponymes associés à 2451 individus<sup>9</sup>. Durant des siècles, il a semblé naturel de baptiser du nom de ces médecins les parties du corps (*Fallopian tubes* d'après Gabriello Fallopio [1523-1562], anatomiste italien) ; les maladies (*Crohn disease* d'après le gastro-entérologue américain Burrill Bernard Crohn [1884-1983]) ou syndromes (*Marfan syndrome* d'après Antoine Marfan [1858-1942], pédiatre français) ; leurs agents (*Yersinia pestis*, le bacille de la peste, qui doit son nom au médecin franco-suisse Alexandre Yersin [1863-1943]) et leurs vaccins (*BCG* d'après Calmette et Guérin) ; les techniques chirurgicales (*Pfannenstiel incision*, une incision transversale pratiquée lors d'une césarienne et inventée par Hermann Johannes Pfannenstiel [1862–1909], gynécologue allemand) et leurs outils (*Swan-Ganz catheter*, qui tire son nom des deux médecins américains Jeremy Swan et William Ganz, qui l'ont inventé au début des années 1970) ; ou encore les tests (*Apgar score* d'après Virginia Apgar [1909-1974], une anesthésiste américaine, que l'on a rétroacronymisé en *Appearance, Pulse, Grimace, Activity, Respiration* comme moyen mnémotechnique puis en *American Paediatric Gross Assessment Record*).

L'anthroponyme n'est pas toujours le nom du découvreur. Il peut s'agir du patient : *Christmas' disease* de Stephen Christmas, ou d'une toute autre personne : il y a quelques années, un chercheur de Chicago a nommé une nouvelle salmonelle *Salmonella mjordan* d'après le joueur de basket-ball américain Michael Jordan.

---

<sup>9</sup> Source : [whonamedit.com](http://whonamedit.com). Consulté le 09/01/2019

Bien qu'ils soient plus courts que les orthonymes (*Chagas disease* versus *American trypanosomiasis*) et qu'ils aient le mérite de rendre hommage aux grands médecins, les anthroponymes présentent des inconvénients. En effet, ils ne disent rien sur l'étiologie de la maladie, sur l'emplacement de la partie anatomique ou sur la nature de l'intervention. Certains leur reprochent de faire honneur à des médecins dont les pratiques ont été contestables voire meurtrières : Hans Reiter (*Reiter syndrome*) et Friedrich Wegener (*Wegener granulomatosis*) étaient des médecins nazis (Woywodt 2007, 424). D'autres les critiquent parce qu'ils sont parfois usurpés : Georges Huntington (1850-1916), qui a donné son nom à la chorée éponyme, ne fut pas le premier à décrire la maladie (Turnpenney & Smith 2003, 153). Par ailleurs, certains anthroponymes varient en fonction des pays : la maladie de Charcot, du nom de Jean-Martin Charcot (1825-1893), en France, est nommée *Lou Gehrig's disease*, du nom du célèbre joueur de baseball américain chez qui elle fut diagnostiquée en 1939, aux États-Unis (Faure, 2012 : 36). Enfin, il peut y avoir une confusion lorsque le nom propre est également un nom commun : *Down*<sup>10</sup> *syndrome* ; *Legg*<sup>11</sup> *disease* ; *Little*<sup>12</sup> *area*.

## 2.2. Les orthonymes et la culture de la scientificité

La volonté des scientifiques d'abandonner les anthroponymes est un phénomène renforcé par le fait que, depuis la fin de la Seconde Guerre mondiale, ce sont les instituts de recherche et les organismes gouvernementaux qui dominent la médecine, et par conséquent, le chercheur individuel se voit progressivement remplacé par des équipes.

Avec les progrès scientifiques qui s'accompagnent d'une forte tendance à vouloir que les dénominations soient aussi précises que possible, la langue médicale remplace donc progressivement ses anthroponymes par des orthonymes (*primary adrenal insufficiency* à la place de *Addison disease* et *uterine tubes* à la place de *Fallopian tubes*), qui vont donner des informations sur la maladie ou l'organe (site, cause, symptôme, etc.).

Pendant, à cause de leur longueur et des difficultés à les mémoriser, ces orthonymes prennent rapidement la forme d'un sigle (*COPD* pour *Chronic Obstructive Pulmonary Disease*) ou d'un acronyme (*SARS* pour *Severe Acute Respiratory Syndrome*). Cette tendance porte un nom : *acronymitis* (Jaffe 1990), une maladie contre laquelle un nombre grandissant

<sup>10</sup> Le mot *down* signifie 'en bas' en anglais.

<sup>11</sup> Le mot *leg* signifie 'jambe' en anglais.

<sup>12</sup> Le mot *little* signifie 'petit' en anglais.

de médecins s'insurgent (Baue 2012). En effet, certains orthonymes sont perçus comme péjoratifs et dénués de toute considération pour la sensibilité du patient : *CATCH22*<sup>13</sup> pour *C*ardiac *d*efects *a*bnormal *f*acies *T*hymic *h*ypoplasia *C*left *p*alate *H*ypocalcaemia *r*esulting *f*rom *c*hromosome *22q11* *d*eletions au lieu de *D*iGeorge syndrome (Tumpenny & Smith 2003).

Cette tendance concerne tous les domaines de la médecine : les maladies (*GERD* pour *G*astro-*E*sophageal *R*eflux *D*isease), les molécules (*ICAM* pour *I*ntra*C*ellular *A*dhesion *M*olecule), les tests (*FAST* pour *F*ocused *A*ssessment with *S*onography for *T*rauma), les interventions (*TURP* pour *T*rans*U*rethral *R*esection of the *P*rostate), les services hospitaliers (*NICU* pour *N*eonatal *I*ntensive *C*are *U*nit) et même les essais cliniques (*JUPITER* pour *t*he *J*ustification for the *U*se of *S*tatins in *P*revention: *a*n *I*ntervention *T*rial *E*valuating *R*osuvastatin) pour lesquels les chercheurs rivalisent d'imagination : *BATMAN* pour *B*isphosphonate and *A*nastrozole *T*rial - *B*one *M*aintenance *A*lgorithm *A*ssessment (Pottegård 2014).

### 2.3. Les toponymes et la culture du voyage

La toponymie est utilisée surtout lorsqu'il s'agit de dénommer des virus émergents : *Ebola* d'après une rivière en RDC, et *Lujjo*, fusion de Lusaka en Zambie et de Johannesburg en Afrique du Sud. Cette tendance est relativement ancienne puisque le nématode nommé *Guinea worm* a été découvert sur les côtes de Guinée au XVII<sup>ème</sup> siècle. L'Australie a connu une épidémie de *Ross river fever* en Nouvelle Galles du Sud en 1928 (cette maladie porte le nom de la rivière autour de laquelle se développent les moustiques porteurs de l'arbovirus). Le virus *West Nile*, qui sévit aux Etats-Unis depuis 1999, a été découvert en Ouganda en 1937. En 1940 a été découverte à Omsk en Russie une borréliose nommée *Omsk hemorrhagic fever*.

Depuis, nombre de virus portent le nom de l'endroit où ils ont sévi ou été isolés pour la première fois : Lacrosse dans le Wisconsin en 1963, Marburg en Allemagne en 1967, Lassa au Nigéria en 1969, etc. On trouve quelques rares toponymes sous forme dérivée : le terme *tular(a)emia* vient de la ville de Tulare en Californie. La rubéole est appelée *German measles* car elle a été décrite par le médecin allemand Friedrich Hofmann, en 1740.

<sup>13</sup> *Catch-22* désigne une situation sans issue. Le terme provient du titre du roman de l'écrivain américain Joseph Heller paru en 1961.

Durant la Première Guerre mondiale, les Américains ont tenté, en vain, de renommer la maladie *Liberty measles*<sup>14</sup>.

Pendant, cette mode n'est pas du goût de tout le monde puisque les Mexicains se sont rapidement élevés contre le terme *Mexican flu*, né à la suite de la pandémie de grippe H1N1 de 2009, qui a été renommée *swine flu*<sup>15</sup> puis *H1N1 influenza* sous la pression de l'industrie porcine. Plus récemment, certains pays du Moyen-Orient n'ont pas apprécié qu'une maladie causée par un coronavirus soit nommée *Middle East respiratory syndrome* ou *MERS*.

Par contre, les habitants de la ville de Lyme dans le Connecticut à laquelle on doit le nom de la borréliose *Lyme disease* sont ravis de ce toponyme à tel point qu'ils ont placé la tique au cœur de leur communication (Crouch 2015).

Pour la première fois, en mars 2015, l'OMS a publié des recommandations<sup>16</sup> concernant la dénomination des nouvelles maladies. Sont désormais interdits les noms de lieux géographiques, de personnes, de professions, d'animaux et de nourriture, ainsi que les mots renvoyant à des références culturelles ou susceptibles de faire naître un sentiment de peur (*unknown, fatal et epidemic*). Pour autant, nombre de maladies et de virus émergents sont encore nommés d'après un lieu ou une personne (Faure 2017).

## 2.4. Les syndromes culturels

Il nous a semblé important de mentionner les syndromes dits "culturels" dans un article qui traite du lien étroit entre terminologie et culture(s). Nous les avons classés par thématique. Tous ces termes sont présents dans le dictionnaire médical en ligne *The FreeDictionary*<sup>17</sup>.

### Culinaire :

*Chinese restaurant syndrome* désigne une intoxication au glutamate, un exhausteur de goût très utilisé dans la cuisine chinoise.

*Steakhouse syndrome* est une obstruction œsophagienne complète due à un bolus alimentaire excessif.

*Sushi syncope* est un malaise soudain causé par l'ingestion de wasabi, un condiment japonais qui accompagne les sushis.

<sup>14</sup> Voir Ford, Nancy Gentile. 2002. *Issues of War and Peace*. Greenwood Publishing Group.

<sup>15</sup> Le mot *swine* signifie 'porc' en anglais.

<sup>16</sup> [http://apps.who.int/iris/bitstream/10665/163636/1/WHO\\_HSE\\_FOS\\_15.1\\_eng.pdf?ua=1](http://apps.who.int/iris/bitstream/10665/163636/1/WHO_HSE_FOS_15.1_eng.pdf?ua=1). Consulté le 09/01/2019.

<sup>17</sup> <https://medical-dictionary.thefreedictionary.com>. Consulté le 16/01/2019.



*Calvarial doughnut lesions-bone fragility syndrome* se caractérise par des lésions en forme de beignets situées au niveau des os de la voûte crânienne, le calvarium.

*Maple syrup urine syndrome* ('la maladie du sirop d'érable' ou "leucinose") est une maladie génétique caractérisée par l'accumulation de leucine dans l'organisme qui lorsqu'elle est éliminée, confère à l'urine une odeur typique de sirop d'érable.

*Coca-cola coloured urine* désigne des urines de couleur sombre, indicatives d'une rhabdomyolyse, une destruction des cellules musculaires squelettiques.

### **Mode de vie :**

*Monday death* est la mort brutale chez des ouvriers travaillant dans des usines de dynamite et déclenchée par une réexposition à la nitroglycérine et à l'éthylène glycol après un weekend de sevrage.

*Holiday heart syndrome* est une anomalie du rythme cardiaque due à une consommation excessive d'alcool, assez courante chez la jeunesse durant les périodes de vacances.

*Saturday night palsy* est une atteinte du nerf radial souvent déclenchée par une compression lors de phases de sommeil. On parle également de *honeymoon palsy*.

*Yuppie flu de Young Urban Professionals* désigne le syndrome de fatigue chronique et fait référence aux jeunes cadres dynamiques et ambitieux aux États-Unis.

Il en existe bien d'autres mais ces quelques exemples montrent à quel point la terminologie relative à la maladie prend sa source dans la culture et les modes de vie.

## **2.5. L'argot et la culture de l'humour**

Parce que la mort est omniprésente dans la pratique médicale, les médecins ont besoin de se protéger en prenant de la distance. Un des moyens est l'humour dont la fonction cathartique est essentielle pour pouvoir continuer à gérer le stress inhérent à la profession médicale.

Pour créer des lexies humoristiques, la langue médicale anglaise recourt à des procédés analogiques tels que la métonymie (*an axe*) 'hache' désigne un chirurgien) et la métaphore (*boneheads* 'crétins'<sup>18</sup> et *cavemen* 'hommes des cavernes' désignent les chirurgiens orthopédiques<sup>19</sup>).

<sup>18</sup> Le mot *bone* signifie 'os'. Il s'agit d'un jeu de mot.

<sup>19</sup> Les chirurgiens orthopédiques sont souvent déconsidérés par les autres spécialistes.

On trouve également des références à la culture télévisuelle et cinématographique : *Blinky the Fish* (le poisson aux yeux globuleux de la série *The Simpsons*) désigne le radiologue, *a Jack Bauer* (le héros de la série 24) est un médecin qui travaille 24h sur 24h, *a Hasselhoff* (d'après David Hasselhoff, l'acteur de la série *Baywatch*) est un patient qui se présente aux Urgences avec une explication bizarre, *a Bunny Boiler* (d'après le personnage du film *Fatal Attraction*) est une patiente obsessionnelle instable, et *a Hollywood code* est une fausse tentative de réanimation pour laisser croire à la famille du patient que tout a été fait pour le sauver.

Les médecins hospitaliers utilisent, en outre, quelques protologismes tout à fait intéressants : *fascinoma* désigne une maladie dont l'intérêt mérite une publication et *fascinectomy* décrit une intervention longue et passionnante ; *administrivia* désigne la paperasserie administrative qui empêche les médecins de faire leur travail ; *a babygram* désigne l'examen radiologique d'un nouveau-né ; *hypoxanaxemia* renvoie à l'anxiété (déficit de Xanax<sup>®</sup>, un anxiolytique) et *acute hyponicotemia* au syndrome de manque du fumeur.

L'analyse des termes argotiques utilisés par les médecins anglo-saxons – généralement hospitaliers – permet de dégager plusieurs cibles : les collègues issus d'autres spécialités – avec une prédilection pour les chirurgiens, les orthopédistes et les urgentistes –, les patientes obèses (*whales* 'baleines'), les patients âgés (*walkers* 'marcheurs', un terme emprunté à la série *The Walking Dead*), les patients présentant un trouble psychiatrique (*nuts* 'cinglé', *wacko* 'taré' et JPN pour *Just Plain Nuts* 'juste complètement cinglé'), les patients SDF (*curly toe* lit. 'orteil recourbé', une métonymie qui fait référence à la forme incurvée des ongles des pieds des personnes sans domicile) et ceux qui reviennent trop souvent (*frequent flyers* en référence aux programmes de fidélisation des compagnies aériennes), et la mort (*discharged to God/heaven*, *boxed* 'mis en boîte', ECU pour *Eternal Care Unit*, etc.) (Goldman 2014).

### 3. La langue de la médecine, une langue métaphorique et métonymique

Pour George Lakoff et Mark Johnson (2003), la métaphore et la métonymie sont des procédés tout à fait différents : « Metaphor is principally a way of conceiving of one thing in terms of another, and its primary function is understanding. Metonymy, on the other hand, has primarily a referential function, that is, it allows us to use one entity to stand for another ». Autrement dit, la métaphore est une façon de concevoir quelque chose en termes d'autre chose et la métonymie utilise une entité



pour en représenter une autre. D'après George Lakoff, nous utilisons les métaphores de façon intuitive pour comprendre les concepts abstraits. Cela signifie que si les métaphores structurent notre pensée en fonction de l'environnement physico-chimique et de la culture, « chaque langue va pouvoir découper le monde de différentes manières<sup>20</sup> » (p.204).

En médecine, la métaphore et l'analogie sont omniprésentes. Dans une étude publiée en 2010 (Casarett et al.) sur la capacité des métaphores et des analogies à améliorer la communication avec des patients atteints d'un cancer, des chercheurs ont identifié, sur un échantillon de 101 conversations<sup>21</sup> entre 52 oncologues et 94 patients, 193 métaphores et 75 analogies. Cette étude a conclu que les oncologues qui avaient recours à la métaphore étaient considérés comme de meilleurs communicants par leurs patients.

### 3.1. La guerre

Les métaphores militaires auraient été introduites par le médecin Thomas Sydenham au XVII<sup>ème</sup> siècle. Mais on en trouve déjà des traces dans les écrits de John Donne en 1627. En 1864, Louis Pasteur a présenté sa nouvelle théorie microbienne à l'aide de la métaphore de l'invasion. En 1971, le Président américain Richard Nixon les reprend dans son allocution : “mobilized the country's resources to make the conquest of cancer a national crusade”<sup>22</sup> (Lane et al. 2013).

La métaphore de la guerre est omniprésente dans la terminologie médicale actuelle. Voici quelques exemples empruntés au médecin Paul Hodgkin (1985, 1820) :

It's an overwhelming infection; she's got an infiltrating carcinoma; the body's defences; he's having a heart attack; killer T cells; we must treat him aggressively and use everything in [the] therapeutic armamentarium; we've wiped out smallpox; go to casualty and the house officer will deal with you.

Au-delà des termes eux-mêmes, c'est toute une vision de la médecine qui est façonnée par cette métaphore. La maladie est toujours vécue comme une agression, le refus d'un traitement comme une capitulation et la mort comme une défaite. Si certains patients peuvent tirer profit de cette métaphore en décidant de se battre avec les armes que le médecin met à leur

<sup>20</sup> “Each language may carve up the world in different ways”.

<sup>21</sup> Ni la longueur des conversations ni le nombre de mots ne sont mentionnés dans l'étude.

<sup>22</sup> National Cancer Institute. The 1971 National Cancer Act. <http://www.cancer.gov/newscenter/1971-nca>. Consulté le 09/01/2019.

portée (antibiotiques, antiviraux, chimiothérapie, etc.), d'autres peuvent, au contraire, souffrir de cette vision imposée. Certaines études démontrent en effet que l'utilisation de ces métaphores se révèle, auprès de patients cancéreux, contreproductive et anxiogène car elle les force à taire leur douleur psychique :

However, for patients, these responses meant resisting the expression of emotional distress rather than the disease. By encouragement to 'fight' and 'be positive', clinicians may therefore collude with patients' emotional suppression (Byrne et al. 2002).

D'autres travaux tendent à prouver qu'une telle vision pousse les patients à développer un sentiment de rejet vis-à-vis de leur maladie alors que la notion d'acceptation est importante pour le processus de guérison (Sontag 1990). Mais il semblerait que cette métaphore ait encore de l'avenir puisqu'on la trouve dans des domaines récents de la médecine comme la génomique : *shotgun sequencing* ou encore *BLAST* pour *Basic Local Alignment Search Tool*.

### 3.2. L'économie de marché

L'entreprise a toujours occupé une place importante en médecine, ne serait-ce que parce qu'elle fabrique des outils de diagnostic, des traitements et des instruments chirurgicaux et orthopédiques qui lui sont essentiels. D'ailleurs, l'utilisation des noms de marques pour renvoyer à des instruments ou dispositifs médicaux est un phénomène métonymique ancien. Ainsi, *gurney* 'brancard' vient de la marque d'une calèche inventé par J. Theodore Gurney en 1883 aux États-Unis, et le nom de l'anticoagulant *warfarin* a été construit à partir de *Wisconsin Alumni Research Foundation* et du suffixe *-arin* de *coumarin* en 1950. Mais cette tendance est en pleine essor : *a Band-aid* (un pansement adhésif), *an Aircast* (une orthèse), *a Clear Blue* (un test de grossesse), *a Steri-Strip* (une bandelette qui permet une suture cutanée), *an Ambu bag* (un ballon d'insufflation utilisé en réanimation) et *un AmniHook* (un crochet qui permet de percer la membrane amniotique).

Parce que les dépenses de soins pèsent de plus en plus lourd sur les budgets des différents états, on observe l'émergence d'une terminologie empruntée à l'économie et au commerce : *healthcare provider*, *healthcare consumer*, *doctor's business*, *care management*, *care package* (l'ensemble des services pour répondre aux besoins d'un patient), *aged care assessment team*, etc.

En 1982, Rashi Fein alertait déjà sur les effets négatifs de cette tendance dont il disait qu'elle traduisait un changement profond d'attitude de la part du médecin et avait un impact sur la qualité de la relation médecin-patient : "In speaking the new language, doctors have adopted the attitudes and methodology of economics — a narrow economics that emphasizes efficiency more than equity". Il critiquait notamment la notion de *compliance*, qui ne respectait pas la liberté du patient.

Effectivement, aux États-Unis, désormais, les professionnels de santé sont obligés de raisonner en termes de *quality-adjusted life expectancy* (le rapport entre le temps restant à vivre et l'utilité de donner un traitement), de *standard gamble* (le patient se voit offrir deux alternatives thérapeutiques – la première le maintient dans son état actuel et la deuxième a deux issues : le retour à la santé et la mort), de *productivity costs* (les coûts de la maladie d'un agent pour son employeur), de *dominance* (qualité d'un traitement qui est à la fois plus efficace et meilleur marché que les autres), de *disease burden* (le coût de la maladie pour la société), d'*optimal ageing* (la promotion d'activités et de soins afin de garantir un vieillissement en bonne santé), ou encore de *clinical governance* (le cadre dans lequel les organismes de soins doivent répondre de la qualité de la prise en charge). Si un patient souffre de douleur chronique, on lui fait signer un *pain treatment agreement* (PTA)<sup>23</sup> ou *pain contract*, un contrat qui l'engage à suivre son traitement (on parle de *compliance* 'observance' et de *adherence* 'adhésion'<sup>24</sup>). D'ailleurs, cette tendance pénètre même l'argot de l'hôpital : *a wallet biopsy* désigne l'étape qui consiste à vérifier la solidité financière d'un patient avant de prescrire des examens coûteux.

Dans cette même veine, on observe également l'entrée de la terminologie de l'épidémiologie et des statistiques dans la langue médicale : *morbidity*, *cost-benefit*, *risk assessment*, *efficiency*, *effectiveness*, *clinical significance*, *cohort*, *community-based*, etc.

### 3.3. L'écologie

Face à l'omniprésence des métaphores guerrières et économiques, George Annas (1995) propose d'utiliser, au travers de termes tels que *sustainable*, *balance*, *responsibility* et *conservation*, la métaphore de l'écologie, qui effectivement correspond plus à l'un des paradigmes scientifiques médicaux actuels, à savoir celui de la génomique, qui

<sup>23</sup> Il concerne surtout les traitements à base d'opiacés.

<sup>24</sup> Les médecins français utilisent l'anglicisme *adherence*.

considère le corps humain comme un écosystème et la maladie comme un déséquilibre au sein de cet écosystème.

La métaphore de l'écologie n'est pas nouvelle. Au début du XVIII<sup>ème</sup> siècle, James Hutton, médecin et géologue écossais, se faisait l'avocat d'une médecine qui tiendrait compte de l'environnement et qu'il nommait *planetary medicine*, (Stewart 2009, 72). D'autre part, l'analogie avec la nature est présente dans la terminologie anatomique (*bronchial tree* 'l'arbre bronchique', *biliary tree* 'arbre biliaire', *taste buds* 'les bourgeons gustatifs', etc.) puisqu'elle est la traduction, via le latin, des termes grecs dans lesquels la faune et la flore étaient déjà une source d'inspiration : *vermis* (en latin : 'ver'), *cochlea* (en latin : 'coquille d'escargot'), *muscle* (du latin *musculus* 'petite souris'), *acinus* (en latin : 'grain de raisin'), *bulb* (du latin : *bulbus* 'oignon'), l'uvule (du latin *uvula* 'petit raisin'), *cortex* (en latin : 'écorce') et *gland* (du latin *glans* 'gland') (Faure 2012).

●utre favoriser une médecine holistique dans laquelle l'être humain se voit replacé dans son environnement naturel, les métaphores écologiques nous semblent s'imposer dans le paradigme médical actuel pour trois raisons dont la première se trouve dans les questions de santé publique actuelles qui dépendent de l'environnement telles que l'infertilité due aux perturbateurs endocriniens, les virus émergents souvent liés à la déforestation, certains cancers (par exemple, les cancers dus à l'utilisation de pesticides dans l'agriculture) et les maladies respiratoires étroitement corrélées au taux de pollution atmosphérique. La prise de conscience récente du lien entre les activités de l'homme sur son environnement et ses problèmes de santé tend à impulser un nouveau tournant à la médecine. La deuxième raison est la place grandissante de la recherche clinique dans la médecine actuelle fondée sur les preuves (*evidence-based medicine*), une recherche clinique majoritairement épidémiologique et de fait fortement environnementale. Les progrès en microbiologie et notamment les travaux sur l'interaction biologique ont imposé la métaphore de l'écologie dans le domaine de la pathogenèse : *saprophyte / commensal flora*, *intestinal microbiome*, etc. La résurgence de la théorie métabolique du cancer<sup>25</sup>, une approche, selon laquelle la cause initiale du cancer serait la fermentation du glucose dans la cellule, fait, elle aussi, la place belle à l'environnement.

Au-delà des termes, la métaphore de l'écologie fait apparaître une nouvelle vision du corps humain considéré désormais comme un écosystème, de la maladie vue comme un déséquilibre, des traitements qui, à l'instar des anticorps monoclonaux, viennent en soutien, et la chirurgie, de plus en plus conservatrice. L'utilisation de la métaphore de l'écologie

---

<sup>25</sup> ●n la doit au biochimiste allemand ●tto Warburg, qui l'a proposée en 1924.

dans la terminologie médicale traduit donc aussi une volonté de la part des médecins – et des patients – de réviser un autre des paradigmes actuels, à savoir celui d'une médecine invasive et déshumanisée.

## Conclusion – La révolution de la médecine 5P

Nous avons démontré que la terminologie médicale anglaise est motivée d'une part par la nature et les contraintes liées à la pratique de la médecine occidentale (la concision et le secret médical) et d'autre part par la culture anglo-saxonne (le mérite, la scientificité, le voyage, l'humour, la nourriture et le mode de vie) qui la font évoluer. À chaque contrainte et à chaque culture ou presque, nous avons pu faire correspondre un procédé linguistique (la troncation et l'hypallage pour la nécessité d'être concis, l'euphémisme pour le besoin d'être discret, l'anthroponymie pour la culture du mérite, l'orthonymie pour celle de la scientificité, ou encore la toponymie pour celle du voyage). Nous avons donné quelques exemples de syndromes culturels ainsi que de termes argotiques qui trouvaient leur origine dans la culture culinaire, télévisuelle et cinématographique. Nous avons également mis au jour des mutations métaphoriques (de la guerre à l'économie de marché et à l'écologie) qui traduisent des changements profonds dans la façon d'envisager la maladie et sa prise en charge.

Ces dernières années, on assiste à l'émergence de nouvelles disciplines telles que la génomique et la médecine translationnelle<sup>26</sup>, de nouvelles spécialités telles que l'immuno-oncologie, d'approches thérapeutiques innovantes comme les biothérapies parmi lesquelles la thérapie cellulaire à partir de cellules souches, et à l'essor des biotechnologies dont les prothèses bioniques sont un exemple. Toutes ces avancées scientifiques et technologiques s'inscrivent dans une révolution médicale : la médecine 5P (Prédictive [utilisation de la génétique pour prévoir la survenue de telle ou telle maladie], Préventive [grâce aux nanotechnologies], Personnalisée [traiter chaque patient de façon individualisée en fonction de ses spécificités génétiques et environnementales], Participative ou "médecine 2.0" [partage de l'information entre patients et professionnels de santé] et fondée sur des Preuves [essais cliniques et études épidémiologiques]).

Cette révolution s'accompagne de l'émergence d'une nouvelle terminologie construite à partir de procédés linguistiques tels que la composition (*care customization* et *systems biology*), la préfixation et suffixation (*epigenomics*, *pathogenomics*, *psychogenomics*, etc.), l'holonymie

---

<sup>26</sup> La médecine translationnelle fait le pont entre la recherche fondamentale en laboratoire et la recherche clinique appliquée au patient.



(*omics*), l'amalgame (*e-patient* et *mhealth* de *mobile health*), et l'élargissement sémantique (*empowerment* 'capacitation du patient'). Reste à présent à effectuer une veille terminologique métaphorique (Resche 2002) afin de voir dans quelle direction ce nouveau paradigme nous mène.

## Références bibliographiques

- Annas, George. 1995. "Reframing the debate on health care reform by replacing our metaphors." *New England Journal of Medicine*, 332: 744–747.
- Baue, Arthur. 2002. "It's Acronymia All Over Again: With Due Reference to Yogi Berra." *Archives of Surgery* 137, No. 4: 486-9.
- Byrne, Angela, John Ellershaw, Christopher Holcombe & Peter Salmon. 2002. "Patients' experience of cancer: evidence of the role of 'fighting' in collusive clinical communication." *Patient Education and Counseling*, 48(1): 15-21.
- Casarett, David, Amy Pickard, Jessica Fishman, Steward Alexander, Robert Arnold, Kathryn Pollak & James Tulsy. 2010. "Can metaphors and analogies improve communication with seriously ill patients?" *Journal of Palliative Medicine*, 13(3): 255-60.
- Crouch, Lizzie. 2015. "The twists and turns of naming diseases." <http://www.bbc.com/news/health-34913764>
- Dirckx, John. 1983 [1976]. *The language of medicine, its evolution, structure and dynamics*. New York: Praeger.
- Faure, Pascaline. 2017. "De Mexican flu à A(H1N1)pdm09 : raisons et pertinence de la tentative de normalisation de l'OMS dans la dénomination des nouvelles maladies infectieuses humaines en anglais." *ASp* 72: 69-81.
- Faure, Pascaline. 2012. *L'anglais médical et le français médical : analyse linguistico-culturelle et modélisations didactiques*. Paris : Editions des Archives Contemporaines.
- Fein, Rashi. 1982. "What Is Wrong with the Language of Medicine?" *New England Journal of Medicine*, 306: 863-864.
- Goldman, Brian. 2014. *The Secret Language of Doctors*. Chicago: Triumph Books.
- Hamburger, Jean. 1982. *Introduction au langage de la médecine*. Paris: Flammarion.
- Hodgkin, Paul. 1985. "Medicine is war: and other medical metaphors." *British Medical Journal*, 291: 1820-1821.
- Jaffe, Bernard. 1990. "Acronymitis." *Surgical Rounds*: 11-12.

- Lakoff, George & Mark Johnson. 2003 [1980]. *Metaphors we live by*. Chicago: The University of Chicago Press.
- Lane, Heather, SueAnne McLachlan & Jennifer Philip. 2013. The war against dementia: are we battle weary yet? *Age and Ageing*: 1-3.
- Pottegård, Anton. 2014. "Search for humorous and Extravagant acronyms and Thoroughly Inappropriate names For Important Clinical trials (SCIENTIFIC): qualitative and quantitative systematic study." *British Medical Journal*, 349.  
<http://www.bmj.com/content/349/bmj.g7092>.
- Resche, Catherine. 2002. "La métaphore en langue spécialisée, entre médiation et contradiction : étude d'une mutation métaphorique en anglais économique." *ASp*, 35-36: 103-119.
- Sontag, Susan. 1990 [1978]. *Illness as Metaphor*. New York: Farrar, Straus and Giroux.
- Stewart, William. 2009. *Deep Medicine: Harnessing the Source of Your Healing Power*. Oakland: New Harbinger Publication & Noetic Books.
- Turnpenny, Peter & Ron Smith. 2003. "Of eponyms, acronyms and ... orthonyms." *Nature Reviews Genetics* 4: 152-156.
- Woywodt, Alexander. 2007. Should eponyms be abandoned?" *British Medical Journal*, 335: 424-425.



**PART II:**  
**DIACHRONY AND SCIENCE**



# CHAPTER FOUR

## DIACHRONIC RESEARCH: A METHODOLOGY FOR REVEALING A DISCIPLINE'S IMPLICIT FRAME OF INDEXICALITY

### DACIA DRESSEN-HAMMOUDA

#### 1. Introduction

Diachronic approaches to text analysis have given rise to a wide range of methods, including historical discourse analysis (Berkenkotter, 2009), historical pragmatics (Jucker, 1995), historical sociopragmatics (Culpepper, 2009), historical speech acts (Jucker & Taavitsainen, 2008), and more recently, diachronic English for Specific Purposes (Alonso-Almeida & Marrero-Morales, 2011; Banks, 2012). Such approaches have long been influential in promoting the idea that socio-historical context is important to consider when describing the linguistic and rhetorical features of specialized writing in English for specific, academic and research purposes. Other noteworthy contributions include Bazerman's (1988) study on the development of writing in physics, Salager-Meyer's (1999) study on changes in referential behavior in medical research writing, Atkinson (1999) and Salager-Meyer's (1994) studies of the evolution of medical written discourse, and Banks's (2008) analysis of the development of English scientific rhetoric across specific journals. These studies have been instrumental in helping frame our understanding of scientific writing. They have also proposed methods that prove useful for examining the harder-to-see features of specialized writing, such as its implicit, "insider" meaning. Not having access to this implicit content can impede less experienced or non-native writers' ability to gain a foothold in their respective fields. As argued elsewhere (Dressen-Hammouda, 2008), becoming a proficient writer in a specialized area does not happen just because one has been exposed to the relevant features of specialized genres in the ESP classroom.

Rather, learners must also incorporate in-depth contextual knowledge about the socio-historically situated processes by which experts in their fields produce specialized writing.

Unquestionably, socio-historical insight into the implicit features of specialized writing can make a considerable contribution to teaching methods, making them both richer and more grounded, thus benefitting ESP teachers and trainee-teachers, and by extension, their students. While ethnographic and qualitative approaches to writing research have become increasingly popular of late as a means of providing such explanations (e.g. Barton and Hamilton, 2000; Lillis & Curry, 2010; Paltridge, Starfield and Tardy, 2016), diachronic analyses also provide important tools, notably by situating current text features within past contexts, and showing how they emerge from those contexts. They can also provide insight into which features become “silenced” over time. For example, while some rhetorical acts, such as making frequent and explicit references to one’s personal research actions, are no longer acceptable in modern scientific writing, diachronic analysis provides evidence that such content may still be present in the writing, albeit tacitly; specialist readers glean it from the text thanks to their ability to “read between the lines” and “fill in the blanks”. Being unable to readily identify and reproduce such tacit knowledge can prove to be a serious obstacle for less experienced or non-native writers. An important task in ESP, EAP and ERP research is therefore to help reveal such hard-to-identify implicit features.

The research described in this chapter shows how a diachronic approach can help reveal the implicit content of disciplinary writing. To illustrate the approach, the analysis will focus on the discipline of geology, examining how geologists, both past and present, tend to both write about, and silence, details of their field research. A helpful concept that can also be used to frame the analysis of implicit writing features is indexicality (Dressen-Hammouda, 2014). The next section will briefly describe this concept. Drawing on the comparative, diachronic analysis of two corpora, a four-step methodology based on Hymes (1972) is then described. Indexical analysis, the method used, is useful for identifying specific implicit, indexical structures in geology writing by distinguishing between what is “formally possible”, “contextually appropriate”, “possibly implied” and “actually attested” (see Hymes, 1972). In conclusion, it will be argued that diachronic studies are the foundation for carrying out indexical analysis. In effect, by diachronically reconstructing the emergence of disciplinary and professional discourses and practices, indexical analysis provides meaningful insight into the implicit frames of reference that operate within and give meaning to academics’ and professionals’ written discursive practices.

## 2. Background to the study

### 2.1 Field writing in geology

As described in earlier studies, geologists describe their field research using a discipline-specific part-genre called the field account (Dressen, 2003). Rather than standing on its own, the field account is almost always embedded within other subsections of the research article, thus explaining its characterization as a part-genre (Dressen & Swales, 2000; Ayers, 1994). Its primary purpose is to contextualize and describe field research by drawing on the contents of the geologist's field notebook, which contains various field measurements, roadmaps, affective comments, and visuals. For insiders, the field account is a rhetorical co-construction that translates what it means to be a field geologist, using the timely marshalling of conclusive field evidence.

How the field account responds to the discipline's implicit concerns of fieldwork reporting is not immediately obvious to the non-specialist, however. What implicit content might this part-genre be signaling to initiated members of the geological community? How are informed readers able to infer the necessary information? Addressing these questions requires establishing a methodology to identify how inferential structures are indexed in the writing.

### 2.2 Indexicality

The main argument made here is that indexical analysis allows the researcher to identify content in the text, which is typically implied or only tacitly referred to. Such content can be problematic for less experienced or non-native writers. The understanding of indexicality used comes from research in language socialization, such as Bucholtz and Hall (2005) and Ochs (1992), for whom an index is the semiotic link between linguistic form and social meaning. As a linguistic form, it derives its meaning from the context of social interaction. At the same time, it also elicits or points to (i.e., "indexes") that meaning. Thus, syntactic phrasing, lexical expressions, politeness formulas, prosodic features (e.g., intonation), paralinguistic features (e.g., pauses, laughter), turn-taking devices, sudden shifts in register, dialect or style, etc., are all imbued with tacitly understood, context-specific meaning. If individuals do not share the same code of inferences to elicit relevant associations, a risk of misunderstanding arises. The risk becomes even more acute if each party attributes a different social meaning (e.g., high value, low value) to the forms, as could very well be the case for less

experienced or non-native writers. Because indexicality also contributes to the maintenance of social authority, control and evaluation (Blommaert, 2010), identifying and analyzing its forms in specialized writing is important in terms of promoting success and avoiding exclusion for such writers. How indexical analysis can help reveal the emergence of indexicality in disciplinary writing practices is demonstrated below.

### 3. Methodology

Carrying out an indexical analysis requires a mixed research design, combining quantitative and qualitative approaches from linguistics and ethnolinguistics. A four-part methodology was developed for this purpose, and builds on a diachronic approach:

- (1) study a disciplinary or professional community's socio-historical context: its practices, belief system, writings, and traditions;
- (2) identify the community's currently significant behaviors using ethnographic and qualitative approaches: participant observation, in-depth interviews with its members;
- (3) determine possible linguistic forms (indexes) that refer to those practices and behaviors using genre-based corpus analysis: lexicosyntactic analysis, analysis of collocations/phraseology, rhetorical discourse analysis;
- (4) triangulate the data and validate the relationship between (1), (2) and (3) using reception studies, standard deviation analysis, etc.

This remainder of this chapter reports on the results of (1) a diachronic analysis of two corpora of field writing in geology, (2) an interview study of currently significant disciplinary behaviors, and (3) a targeted corpus analysis intended to reveal relevant linguistic indexes. It draws on results of a larger indexical study (Dressen-Hammouda, 2008, 2014).

#### 3.1 Materials: Two corpora of field writing in geology

To reveal some of the indexical structures found in today's field geology writing, two different corpora have been targeted for analysis.

##### *Corpus 1.*

A first corpus of 115 field accounts was used to examine how early field geologists wrote about their field research. The corpus consists of two parts. The first part consists of 105 English-language articles published between

1665 and 1920 in the *Philosophical Transactions of the Royal Society of London*. The articles were chosen because the title referenced both the name of a locale, as well as some geological object (“load stone”, “fossil”, “iron ore”) or process (“volcanic eruption”, “sinking down of hill”). The second part of the corpus, 10 articles in French, shares these attributes. Published in *Les Comptes rendus de l’academie des sciences* between 1890 and 1920, the French-language articles were included to highlight how fieldwork writing directly prior to WWII was shifting in similar ways in both English and non-English speaking countries. The whole corpus was thematically coded for the ways in which geologists make explicit reference to their fieldwork missions.

### *Corpus 2.*

A second corpus of 88 field accounts, published between 1983 and 2009, was used to examine how field geologists write about their field research today (Dressen-Hammouda, 2014). Three subfields in geology (geochemistry, petrology, and structural geology) represented by nine journals were targeted for analysis. Once again, articles that reported on fieldwork were singled out. As a consequence, all articles in the corpus contain evidence of the actual field presence of at least one of its authors, although determining field presence was not always clear-cut. Field presence can be determined with reasonable accuracy by using a combination of textual clues including whether: (1) the region studied is indicated in the title; (2) the authors make reference to their fieldwork mission in the acknowledgments; (3) they cite their own fieldwork in the bibliography; (4) they credit themselves for the geological maps and other visuals; and (5) the article contains a number of explicit linguistic elements (indexes) that point to the author’s presence in the field.

## 3.2 Participants

In order to validate the characteristics of field writing practices and ascertain what implicit information field writers might be communicating, I strove to “get inside” the discipline (Geertz 1973) and socialize in its practices. Much of this socialization resulted from the proximity I have maintained with multiple geologists over several years, by accompanying them during trips into the field or being present at informal discussions outside the laboratory. Another source of information was to have geologists to talk about their field experiences directly. To this end, semi-structured, in-depth interviews were conducted with four field geologists. Three were full-time researchers, and had been on the job for over ten years



at the time of the interviews. Interviews were conducted over an initial period of six months, followed by shorter interviews over the next two years. The fourth geologist was a doctoral student; interviews were carried out over a period of five years. During this time, the dissertation was completed and full-time university employment was obtained.

Each interview was recorded and transcribed in its entirety; results were coded to highlight evidence of what geologists consider contextually appropriate topics to write about.

### 3.3 Analytical procedure: Identifying indexical structures

The main analytical difficulty involved in identifying indexical structures is that, for the most part, they are implicit, and thus not visible. The approach I have used to try and get around this difficulty is to compare and contrast what is *actually said* with what *could have been said* (Huckin 2002). To do so, I have adapted Hymes' (1972) distinction between what is "formally possible", "contextually appropriate", and "actually attested" in a discourse, and applied it to the contrastive analysis of the two different corpora. The analytical procedure used is as follows:

- (1) Using the corpus of early field writing (1665-1920), I identified the **formal discursial possibilities** of the field account part-genre;
- (2) **Contrasting** those features with the corpus of modern field writing (1983-2009), I validated which of those discursial possibilities **have likely become contextually inappropriate** today, i.e., because they are not found in the later corpus;
- (3) Carrying out several long-term, in-depth interviews and a small-scale ethnography (several day-long participation-observations carried out with geologists in the field), I identified **what geologists still consider to be contextually appropriate** today. It is plausible to expect that such content is indexed in field writing;
- (4) Keeping in mind the distinction between past discursial possibilities, current contextual appropriateness and inappropriateness, I then identified what is **actually attested** in modern geological field writing in order identify which linguistic cues may be indexing inferential structures.

## 4. Results

### 4.1 Seeking formal discursual possibilities

Using Gohau (1987) and Ellenberger (1988) to contextualize the emergence of geological practice, and drawing on Rudwick's (1985) description of the discipline's development between 1800-1850, the corpus of early field writing was analyzed, spanning the years 1665 to 1920. The field writing from this time was found to possess the following features:

- nouns denoting geological objects;
- a narrative of exploration providing a chronology of the researcher's actions;
- grammatical subject roles depicting the researcher;
- specific verbs indicating the researcher's physical and intellectual acts;
- references to oneself so as to build credibility and believability;
- details about the dangers or ordeals encountered in the field;
- situated site descriptions;
- drawings of the site.

Four sample texts were selected from Corpus 1 to illustrate these features. The first excerpt, taken from a text published in 1668, recounts the field experiences of an "Ingenious English gentleman" when visiting Mexico:

*Text 1 – 1668:*

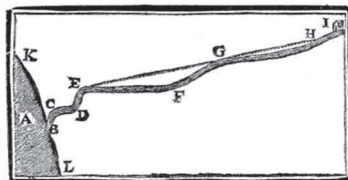
I was once desired to visit a famous Cave there, some Leagues from Mexico on the North-west side of the City beyond the Lake. [...] I rid thither one morning, taking with me one Indian on'y for my guide, with a Tinder-box and a Candle, and some other instruments for my design. I found it situated somewhat high, in a place very convenient for generation of Mettals, but the mouth so barricaded with stones, that both my Indian and I had to work enough to clear the passage for my entrance, which being open'd, I went in with my Candle lightned, but could not make the Indian follow me, being afraid of Spirits and Hobgoblins. The light of the Candle soon discover'd to me on all sides, but especially above my head, a glittering Canopy of the said Mineral Leaves, at which I greedily stretching forth my hand to reach some parcels of it, there fell down presently so great a lump of clotted sand on my head and shoulders, that not only it put out my Candle, but my eyes also."

This first excerpt is a good illustration of the type of field writing common of the time. After situating the site location (“a famous Cave, some Leagues from *Mexico* on the North-west side of the City beyond the Lake”), the writer structures the entire account as a personal narrative of exploration presented chronologically (“I rid thither one morning”, “my Indian and I had to work enough to clear passage”, “I went in with my Candle”). A number of events, unexpected by modern standards of objective scientific reporting, are also raised: “greedily stretching forth [his] hand”, “a [great] lump of clotted sand [fell] on [his] head and shoulders, [putting out both his] Candle, but [his] eyes also.” While in the end, the events recounted appear more amusing than seriously dangerous, there is a noted tendency to construe the field mission as an ordeal.

The second extract, from a text published in 1714 describing land sinking along the coast of Kent, focuses less on the ordeal of fieldwork than on placing the geologist squarely in the field, this time with the purpose of providing a full enough description so as to make his claim appear “true”, and therefore convincing.

*Text 2 – 1714:*

“Sir, I am about to give you the best Account I can of what is remarkable, and known almost to all hereabouts, concerning the pressing forward of the Cliffs, and sinking of the Hills in the Neighbourhood of or Town of *Folkestone*. I begin with giving you a Sketch of the Situation of the Country. [...]



.A. The *Missing-Rock*, about half-way between high and low Water-Mark.

D d d d

B

[...] Now, Sir, let us climb both these cragged Cliffs, and place our selves at the Top of the higher One, at the Point E. And here we are to observe, that (as old Men inform us) upward of forty Years ago, not so much as the Top of *Tarlingham-House* could be discern'd, neither from hence, nor yet a good Distance off at Sea; but it discover'd it self by degrees, till at this Day, not only the whole House, but a great Tract of Land below it, is plainly to be seen [...]. Whatever be the Cause of [the sinking], 'tis not to be doubted but that these Hills are greatly sunk. [...] I thought it proper to give you this account of the Nature of the Earth; and withal to mingle with it the Opinion of the People, that you might perceive they are so far from doubting the Truth of what is abovementioned [...]. I assure my self that I have Credit enough with you to be believed upon my own single Subscription...

The author once again begins by situating the geological location of his observations (“in the Neighbourhood of or Town of *Folkestone*”). To establish the basis for his claims, he includes a sketch and then sets out geological facts and reasoning for his reader. He combines his own description and that made by “village elders” (“as old Men inform us”) to give more credence to his claims, all the while situating his reader within his own projected visualization of the site, as if the reader, too, were present in the field (“Now, Sir, let us climb both these cragged Cliffs, and place our selves at the Top”, “And here we are to observe, that”). In writing the field description, the author strives to describe the field in a way which makes the observations seem not only plausible, but even obvious. In effect, the field description must describe natural reality in a way which indicates its self-evident nature — making it as plain as “the nose on one’s face” (Rudwick, 1985, p. 103). Thus, the writer assumes the veracity of his field description, such that the point he is trying to make is “plainly to be seen” and “not to be doubted”. While he supports the case for the soundness of his account with natural and social proof (the “Nature of the Earth”, “mingl[ed] with the Opinion of the People”), at the same time he hopes his personal standing will allow his reader to take his account at face-value (“I assure my self that I have Credit enough with you to be believed upon my own single Subscription”).

The importance of social recognition in having one’s field observations accepted at face-value took on critical significance during the first part of the 1800s, during the so-called Devonian controversy (Rudwick, 1985). One of the most significant actors of the period, Charles Lyell (a mentor to Darwin), figures in Corpus 1 as author of an article published in 1835, in which he discusses proof for the “gradual rising of land in certain parts of Sweden”. It is noteworthy that despite his obvious notoriety, which he marshals so as to make way for his own viewpoint, Lyell cannot escape the necessity of also having to produce solid field evidence for his claims:

*Text 3 – 1835:*

Notwithstanding the numerous proofs recorded of the change of level [of land in Sweden], and the high authorities who had declared in its favour, I continued, in common with many others, to entertain some doubts respecting the reality of the phenomenon, partly because I suspected it might be explained by reference to more ordinary causes ... and partly because it appeared to me improbable that such great effects of subterranean expansion should take place in countries which, like Sweden and Norway, have been remarkably free within the times of history from violent earthquakes. [...] I am willing, however, to confess, after reviewing all the statements published... that my scepticism appears to have been unwarrantable [...] I propose, therefore, to lay before the Royal Society

**the observations which I made during the summer of 1834, with a view of satisfying myself in regard to the data. [...] Proceeding northwards along the coast of the Baltic, the first place which I visited where any elevation of land is supposed to be going on was Calmar... situated in latitude 56° 41'. [...] When I observed that the base of one of [the towers] rested on the beach only two feet above the level of the water, and when I found that sea-weed had recently been washed up, so as to touch the lowest part of the building, I concluded, at first, that ... we should be compelled to suppose that part of the tower had been originally constructed under water. But on nearer inspection I was led to suspect that this had really been then case, and that the foundation was originally subaqueous. At the height of about two feet above the base of the tower (see sketch, fig. 1), and four feet above the level of the sea, a projecting band of stone ... “**

Text 3 once again shows a typical use of first-person pronouns (“I visited”, “I observed”, “I found”, “I concluded”), situating the author within a step-by-step narrative of his discoveries. A significant amount of textual space is dedicated to his intellectual acts (“I entertained doubts”, “I am willing to confess that my scepticism is unwarrantable”, “I was led to suspect”). At the same time, he neatly locates his site descriptions (“Proceeding northwards along the coast of the Baltic”, “Calmar... situated in latitude 56° 41’”) and relies on description to bolster his claim (“At the height of about two feet...”).

In Text 4, published in French nearly 80 years later, in 1912, the same structures are still apparent: a chronological narrative indicates an unfolding sequence of actions the geologist has undertaken in the field (“While climbing up the bed of the [...] River... I first encountered”, “whose trace I later found”). Like the preceding texts, the narrative is also highly personalized, here so much so that the geological details seem to take a back seat to what happened to the geologist in the field. From a 21<sup>st</sup>-century point of view, we can note with some amusement that the writer actually tells us his fieldwork was “difficult” and, especially, that it was “dangerous”:

*Text 4 – 1912:*

**While climbing up the bed of the little Paillacardriver, a task fraught with difficulty and at times with danger, I first encountered a series of lime blocks, whose trace I later found in the deep ravines of Morne Fortuné on the left bank of the Céron river... (Dublanq-Laborde, 1912; translated from original text in French)**

He goes on to explain that fortunately for him, his hosts’ spear-carrying servants had accompanied him, thus protecting him from potential attack by natives.

What is noteworthy in these excerpts are the sheer number of linguistic strategies available to writers of the time for establishing their physical presence in the field. They could, for example, extensively use first person pronouns and a range of verbal strategies to explicitly describe their physical and intellectual activity. Moreover, by describing the surrounding locale they have provided not only more or less explicit instructions on how to get to the site, but also undeniable proof they were in the field (e.g., “near the Puyferret residence”, “the bed of the little Paillacard river”, “the deep ravines of Morne Fortuné”, “the left bank of the Céron river”). A number of such highly textualized indicators of physical field presence allowed these early geologists to explicitly set the basis for an evaluation of their credibility (Dressen, 2003). This, in turn, allowed for them to establish their authority over their territory by showing beyond a doubt that their field observations were “true”.

### Identifying what is contextually *inappropriate*

Of course, a discipline’s formal discursual possibilities change over the years because what is contextually appropriate at one point in time later becomes inappropriate in response to changing disciplinary contexts, social needs and institutional imperatives. What once was sayable is no longer so, even if the topic may continue to hold importance for insiders. The centrality that fieldwork long enjoyed, for example, underwent a significant shift toward the periphery of geological science around 1930-1940. In post-war disciplinary practices, fieldwork was considered far less central as many geologists turned away from merely collecting empirical field evidence to developing newer, more experimental methods.

A comparative study of discursual trends using Corpus 2 (1983-2009) confirms the growing contextual inappropriateness of many of the topics previously seen. Notably, the results show that conventions for reporting moved toward a greater textual discretion on the part of the researcher, accompanied by a marked diminishment of overt authorship. This shift toward greater “authorial invisibility” is a feature similarly noted in other areas of scientific discourse, such as physics (Bazerman 1988) and medical discourse (Salager-Meyer 1999). The following excerpt clearly shows little resemblance to the narratives of discovery and researcher-centered accounts of ordeal and exploit typical of earlier field writing:

In the current (SACS 1980) classification the Bullstroom Formation eruptives occur only below the level of intrusion of the RLS components and comprise about 80 flows of predominantly basaltic andesite composition with minor basalt, dacite and rhyolitic units. Two chemical sub-classes are distinguishable in the intermediate components: a high iron-titanium group



(Fe<sub>2</sub>O<sub>3</sub> ~ 11-15%; TiO<sub>2</sub> 1.2-2.2%) and a more common low-titanium group having TiO<sub>2</sub> and Fe<sub>2</sub>O<sub>3</sub> ~ below 1% and 11% respectively. The Røisberg acid eruptives form the roof of the RLS in the Loskop Dam area where Twist (1985) subdivided the ca. 3.5 km thick volcanic succession into 9 units on the basis of colour, texture, phenocryst content and internal structure. (Harmer & Farrow 1995)

In this excerpt, the grammatical actor-role has effectively shifted away from the researcher toward the geological structures being studied, resulting in the apparent disappearance of the geologist from the field account. The verbs no longer focus on *human* activity, but on *rocks* and *geological structures*. Human researchers appear almost completely absent, despite the obvious necessity of such inherently human activities as observing, mapping, or interpreting geological structures (Dressen & Swales, 2003). Table 1 summarizes the topics that appear to have become contextually inappropriate in today's geological field writing.

Table 1. Contextually *inappropriate* topics in the modern field account

Contextually inappropriate topics	Reasons for the shift
– Writing only about fieldwork	– Less consideration is given to fieldwork because it is no longer the discipline's sole methodology; accordingly, publications report on other methods as well.
– Talking about danger or ordeal	– Narratives of discovery or exploit are no longer acceptable.
– Providing a detailed narrative of the researcher's actions in the field	– Accounts of field research in geology have, like in other sciences, become "detemporalized" (Myers 1990) in order to provide a more neutral and distanced account.
– Explicitly talking about oneself	– It is no longer conventional to talk about oneself. Grammatical agent roles have largely shifted from a focus on the researcher to the object of study; most verbs no longer describe human activity, but rock activity.

### 4.3 Identifying what is contextually appropriate, and possibly implied

Needless to say, it is not because a community no longer writes overtly about certain topics or concerns that those topics have lost all relevance, or



that insiders no longer refer to them at all. Field geology makes an interesting case study in this regard because the fundamentals of doing fieldwork have remained basically unchanged for the past 300 years. The results of participant observations and extensive in-depth interviews have revealed that the problems and concerns of the past continue to be central to the field geologist's experience today. Doing fieldwork requires an acceptance of work as ordeal, i.e., physical exertion, exposure to the elements, sensitive or risky geopolitical contexts, and of being away from home for lengths of time. As shown in Table 2, contextual analysis indicates that these topics are still quite important for field geologists; as such, we can consider them *contextually appropriate*. This content provides a promising avenue of investigation to confirm whether certain indexes tacitly refer to these topics, although they have become inappropriate to write about explicitly.

Table 2. Contextually *appropriate* topics point to possible indexes

- Proving authenticity as a real field geologist
- Proving one does fieldwork
- Showing physical presence in the field
- Demonstrating an ability to see and interpret correctly
- Demonstrating proper disciplinary training
- Demonstrating knowledge about proper terms, what to describe
- Demonstrating an ability to use current frames of interpretation
- Promoting one's own fieldwork

#### 4.4 Identifying what is actually attested

In order to determine how such topics may be indexed in field writing, I re-examined Corpus 2 with this purpose in mind. It was surmised that field geology's "indexes" (Bucholtz & Hall 2005, Ochs 1992) refer both readers and writers to "some aspect of the situation at hand [that] is *presupposed* or even *created*" (Duranti, 2001, p. 458, original emphasis). To identify possible indexes, I applied a *qualitative content* analysis (Salager-Meyer 1999, Huckin 2002), manually examining each article. I focused on:

- Syntactic structures, such as complex noun phrase constructions, verbal adjectives and participles, adverbs, adjectives, and pronominal structures;
- Evaluatives, such as adjectives and adverbs;
- Explicitly geological terms indicating measurements and/or

- specialist terms or concepts; <sup>SEP</sup>
- Metadiscourse statements referring readers to visual representations of the fieldwork, or to prior publications;
  - Sentence-level structures that argue for a particular interpretation of field structures.

Corpus 2 was coded for 13 indexes found to specifically relate to doing fieldwork (Dressen-Hammouda, 2008, 2014). Rather than seek *a priori* categories to compare overall similarities with other corpora of scientific discourse (e.g., Biber & Finnegan, 1989; Hunston & Thompson, 2000; Hyland, 2000), I identified the 13 indexes as *a posteriori* categories in order to target the features distinctly evocative of the “appropriate” content revealed by contextual analysis. Three overarching categories appeared to characterize the rhetorical purpose of the field account: (1) showing that the author was personally invested in the field study (personalization indexes); (2) proving the fieldwork was carried out by the author by giving details only the author could provide (doing-the-work indexes); and (3) demonstrating that the author’s fieldwork is embedded within appropriate disciplinary frames of knowledge and interpretation (disciplinarity indexes). Examples of the 13 indexes, taken from Corpus 2, are given below:

### Personalization indexes

**We have also** observed definite sheath folds in oblique sections...

Because the block is exposed high in the near-vertical wall of a glacial cirque... **we have not been able to** examine it closely.

Kinematic indicators are less clear and abundant in rocks.

The CFS is extremely well exposed along most of its strike, but exceptional exposures occur along the south and east sides of the Sierra Cabrera basement high

It is enigmatic that if the pyroxenite was indeed the earliest intrusive component, it is never seen intruding the mafic fenites that are extremely well-exposed along the...

1. Author pronouns/ possessive adjectives

2. Evaluative adjectives and adverbs showing researcher discernment

3. Interpretive comments about field structures, based on field observation

**Doing-the-work indexes**

**Geological mapping of stretch lineations** (Fig. 7), **shear indicators** (Fig. 8a, b) and **other structures** reveal late ductile deformations...

**No feldspar, magnetite or mica is ever found** in these ijolites.

**Maximum displacements on the mapped faults range from 1.5 m to 150 m, and their trace lengths range from 108 m to 6584 m.**

**The peridotite overlies high-grade gneisses and marbles... along an essentially low-angle brittle thrust marked by extensive brecciation discernible over a distance of up to 100 m away from the context.**

**The most spectacular slickenslide surfaces have been schematically represented in Fig. 2 and they are shown in the photographs of Fig. 3.**

**The Koolen Lake-Lavrentiya Bay region** (Figures 2, 3, and 4) **exposes a sequence of sillimanite-grade to second-sillimanite-grade granitic gneisses, paragneisses, schists, amphibolites, and marbles ...**

*(Author, 1996)*

**Disciplinarity indexes**

**Even in the anomalous Maggia steep zone, this radial pattern of flat lying lineations is not disturbed, the lineations following the strike of the steep zone.**

**Miocene-Recent left-lateral oblique-slip (transpressional) was accommodated along the CFS.**

**... the HTD lineation is flat lying, often pitching gently E, but sometimes even gently W. If the foliation is restored to the horizontal by rotating about the E W strike...**

4. **Nominal and verbal markers of researcher activity**

5. **Metric, angle and direction measurements**

6. **Locational adverbs and prepositions, showing researcher movement in the field**

7. **References to visual field data, such as maps, photos...**

8. **Geographical location of the fieldwork site**

9. **Self-citation of prior fieldwork**

10. **Nominal or adjectival field descriptors**

11. **Indicators of geological age**

12. **Technical verbal adjectives and participles**

Backfolds postdate well developed biotites which have been dated at 15 Ma (Steiger and Bugher 1978, Ramsay in Steck et al. 1979).

13. References to others' fieldwork

Validating the socio-cognitive reality of such indexes is a relevant next step. An unpublished reader-response study (cf. Tardy & Matsuda, 2009) has examined disciplinary specialists' sensitivity to their presence or absence in the field account. Although space constraints do not allow for detailed discussion here, the results strongly suggest that trained field geologists are in fact very sensitive to the indexes because they allow them to assess the author's disciplinary identity, skill level and credibility. The indexes described in this chapter thus appear to provide meaningful information to disciplinary insiders, although they remain mostly invisible to non-specialists to whom they may even convey a completely different meaning.

## 5. Implications and conclusion

This chapter has described how to carry out indexical analysis, a method for unveiling some of the more "occluded" (Swales, 1996) aspects of specialized writing. Its diachronic component shows how the indexes identified in a corpus reflect the historically-situated concerns of disciplinary practice. By combining these results with contextual analysis (in-depth interviews, participant-observation, reader-response studies), indexical analysis substantiates what would otherwise be suppositions about the sorts of information disciplinary writers may find important to tacitly convey in published writing.

In many respects, diachronic analysis is requisite for understanding how indexical structures emerge and call attention to situated meaning. It adds a necessary depth to essential insights brought by other qualitative and ethnographic approaches. In his work on the early development of English scientific rhetoric, Banks (2017) for example has argued that a discipline's historical situation is key for understanding how its rhetorical dynamics arise from specific contexts. By affording a richer and thicker understanding of social experience, diachronic analysis provides an analytical frame well-suited to understanding the text-context relationship. A diachronic view of context inherently seeks to explain why texts look like they do, what particular linguistic choices are available, and what meaning, purpose and use they have. As such, it moves beyond an analysis of situation by paying close attention to how language is longitudinally embedded within the context of culture.

Indexical analysis contributes to this effort by recognizing that indexes are the cornerstone of language and social interaction. As intermediary structures, indexes help bridge the gap between text and context (Lillis, 2008). As indicators of a community's shared ways of knowing, they tie past and present practices. They construct the sociocultural environment that gives rise to the communicative practices which inform specialists' and novices' involvement in that community. They provide the means by which novices are socialized into their communities (Ochs & Schieffelin, 2014), and further contribute to processes of legitimization, control and authority (Blommaert, 2010).

Unlike formalist linguistic studies, indexical analysis does not assume a split between text and context. Because of this premise, indexical analysis "radically challenges the [traditional] dichotomy between language and culture" (Lillis, 2008, p. 381). As such, it proposes a more compelling analytical frame for studying the implicit and tacit features of specialized writing than other commonly used approaches, such as Grice's (1975) theory of implicatures. Largely as a result of Grice's own ambivalence about distinguishing between cognitive and societal approaches to meaning (Cap, 2009), linguistic applications of his theory have remained predominantly limited to a view of universally shared, cognitively-based cooperation principles of communication. To the contrary, the research described in this chapter provides strong evidence for the need to pay close attention to the situated, historical context of writing of each discipline. Although the search for universals often reveals interesting patterns, those patterns, without grounding in a discipline's "substance", run the risk of remaining convenient analytical artefacts.

## References

- Alonso-Almeida, Francisco, and Sandra Marrero-Morales. 2011. "Introduction to the Special Issue 'Diachronic English for Specific Purposes.'" *Revista de Lenguas Para Fines Especificos* 17: 13–22.
- Atkinson, Dwight. 1999. *Scientific Discourse in Sociohistorical Context: The Philosophical Transactions of the Royal Society of London, 1675-1975*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Ayers, Gail. 1994. "Are Abstracting Changing? A Preliminary Investigation through the Analysis of the Short Texts Accompanying the Articles in Nature." Unpublished MA dissertation, UK: University of Birmingham.
- Banks, David. 2008. *The Development of Scientific Writing: Linguistic Features and Historical Context*. London: Equinox.



- . 2017. *The Birth of the Academic Article: Le Journal Des Sçavans and the Philosophical Transactions, 1665-1700*. Equinox: Sheffield, UK.
- Barton, David, Mary Hamilton, and Roz Ivanič, eds. 2000. *Situated Literacies: Reading and Writing in Context*. London, UK: Routledge.
- Bazerman, Charles. 1988. *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*. Madison: University of Wisconsin Press.
- Berkenkotter, Carol. 2009. "A Case for Historical 'Wide-Angle' Genre Analysis: A Personal Retrospective." *Ibérica* 18: 9–22.
- Biber, Douglas, and Edward Finnegan. 1989. "Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect." *Text* 9: 93–124.
- Blommaert, Jan. 2010. *The Sociolinguistics of Globalization*. Cambridge: Cambridge University Press.
- Bucholtz, Mary, and Kira Hall. 2005. "Identity and Interaction: A Sociocultural Linguistic Approach." *Discourse Studies* 7: 585–614.
- Cap, Piotr. 2009. "Implicatures in Discourse Space or Grice's Rationality Reconsidered." *Topics in Linguistics* 4: 8–11.
- Culpeper, Jonathan. 2009. "Historical Sociopragmatics. An Introduction." *Journal of Historical Pragmatics* 10: 179–86.
- Dressen, Dacia. 2003. "Geologists' Implicit Persuasive Strategies and the Construction of Evaluative Evidence." *Journal of English for Academic Purposes* 2: 273–90.
- Dressen, Dacia, and John M. Swales. 2000. "Geological Setting/Cadre Géologique in English and French Petrology Articles: Muted Indications of Explored Places." In *Analysing Professional Genres*, edited by Anna Trosberg, 57–76. Amsterdam: John Benjamins.
- Dressen-Hammouda, Dacia. 2008. "From Novice to Disciplinary Expert: Disciplinary Becoming and Genre Mastery." *English for Specific Purposes* 27: 233–52.
- . 2014. "Measuring the Voice of Disciplinarity in Scientific Writing: A Longitudinal Exploration of Experienced Writers in Geology." *English for Specific Purposes* 34: 14–25.
- Duranti, Alessandro. 2001. "Intentionality." In *Key Terms in Language and Culture*, edited by Alessandro Duranti, 129–31. Oxford: Blackwell Publishing.
- Ellenberger, François. 1988. *Histoire de La Géologie*. Lavoisier: Paris.
- Geertz, Clifford. 1973. *Local Knowledge: Further Essays in Interpretive Anthropology*. New York: Basic Book.
- Gohau, Gabriel. 1987. *Histoire de La Géologie*. Paris: La Découverte.



- Grice, Herbert Paul. 1975. "Logic and Conversation." In *Syntax and Semantics 3: Speech Acts*, edited by P Cole and J Morgan, 41–58. New York: Academic Press.
- Huckin, Thomas. 2002. "Textual Silences and the Discourse of Homelessness." *Discourse & Society* 13: 347–72.
- Hunston, Susan, and Geoffrey Thompson, eds. 2000. *Evaluation in Text: Authorial Stance and the Construction of Discourse*. Oxford: Oxford University Press.
- Hyland, Ken. 2000. *Disciplinary Discourses: Social Interactions in Academic Writing*. Harlow, UK: Longman.
- Hymes, Dell. 1972. "Models of the Interaction of Language and Social Life." In *Directions in Sociolinguistics: The Ethnography of Communication*, edited by John Gumperz and Dell Hymes, 35–71. New York: Holt, Rhinehart & Winston.
- Jucker, Andreas, ed. 1995. *Historical Pragmatics. Pragmatic Developments in the History of English*. Pragmatics & Beyond New Series 35. Amsterdam/Philadelphia: John Benjamins.
- Jucker, Andreas, and Irma Taavitsainen, eds. 2008. *Speech Acts in the History of English*. Pragmatics & Beyond New Series 176. Amsterdam/Philadelphia: John Benjamins.
- Lillis, Theresa. 2008. "Ethnography as Method, Methodology, and 'Deep Theorizing': Closing the Gap between Text and Context in Academic Writing Research." *Written Communication* 25: 353–88.
- Lillis, Theresa, and Mary Jane Curry. 2010. *Academic Writing in a Global Context: The Politics and Practices of Publishing in English*. London: Routledge.
- Myers, Greg. 1990. *Writing Biology: Texts in the Social Construction of Scientific Knowledge*. Madison: University of Wisconsin Press.
- chs, Elinor. 1992. "Indexing Gender." In *Rethinking Context: Language as an Interactive Phenomenon*, edited by Alessandro Duranti and Charles Goodwin, 335–58. Cambridge: Cambridge University Press.
- chs, Elinor, and Bambi B. Schieffelin. 2014. "The Theory of Language Socialization." In *The Handbook of Language Socialization*, edited by Alessandro Duranti, Elinor ●chs, and Bambi B. Schieffelin, 1–21. West Sussex, UK: John Wiley & Sons.
- Paltridge, Brian, Sue Starfield, and Christine M. Tardy. 2016. *Ethnographic Perspectives on Academic Writing*. Oxford: Oxford University Press.
- Rudwick, Martin. 1985. *The Great Devonian Controversy: The Shaping of Scientific Knowledge Among Gentlemanly Specialists*. Chicago: University of Chicago Press.

- Salager-Meyer, Françoise. 1994. "Hedges and Textual Communicative Function in Medical English Written Discourse." *English for Specific Purposes* 13: 149–79.
- . 1999. "Referential Behaviour in Scientific Writing: A Diachronic Study." *English for Specific Purposes* 18: 279–305.
- Swales, John M. 1996. "Occluded Genres in the Academy: The Case of the Submission Letter." In *Academic Writing*, edited by Eija Ventola and Anna Mauranen. Amsterdam/Philadelphia: John Benjamins.
- Tardy, Christine M., and Paul Matsuda. 2009. "The Construction of Author Voice by Editorial Board Members." *Written Communication* 26: 32–52.

CHAPTER FIVE

THE POPULARISATION OF SCIENCE  
IN 19<sup>TH</sup> CENTURY FRANCE:  
THE “BIBLIOTHÈQUE UTILE”  
OR THE USEFUL LIBRARY

JANET ORMROD

**Introduction**

Much has been written about the development of scientific discourse within the broader topic of academic writing, which is also a major topic in Systemic Functional Linguistics (SFL). However, most of these studies have examined texts written by scientists and other professionals, to be read by their peers. Much less attention has been directed towards scientific texts destined for the general public, especially texts from the 19<sup>th</sup> century in France, when the popularisation of science became a major concern of philanthropists in the years before the republican Jules Ferry had laws passed (in 1881 and 1882), making primary education free, compulsory and non-clerical.

My corpus examines texts intended for a lay audience, and compares them to texts on the same topic that appeared in a scientific journal at the same period, in the mid-19<sup>th</sup> century. My study is based on one section from Volume I of the “*Bibliothèque utile résumant ce que chacun doit savoir*” (“*Bibliothèque utile*<sup>1</sup> resuming what everyone must know”), written by Joseph Morand and published in 1859, and one comparable section from an issue of the *Journal des Savants*, treating the same topic: Kepler’s laws of planetary motion, written by Jean-Baptiste Biot, and published 13 years earlier in 1846. My choice of the *Journal des Savants* enables me to compare the characteristics of professional scientific writing in France at

---

<sup>1</sup> The “Useful Library”, my translation

that time with the popularised version. To undertake the analysis, I follow Halliday and Martin's "Writing Science", 1993, and in particular Halliday's chapter 3: *On the language of physical science*, which analyses scientific discourse in English, with its nominalisations, grammatical metaphor and other significant features. In addition to nominalisation and grammatical metaphor, this chapter also examines retrospective labels and interpersonal interaction. Obviously, Halliday was not looking at French, but a lot of the phenomena he examines are also present in scientific writing in French.

My hypothesis when beginning this study was that the popularised science discourse, "low science" text from the *Bibliothèque utile*, intended for the general public would include, for example, fewer nominalisations, less use of grammatical metaphor, and greater interpersonal interaction, compared to the academic scientific discourse, "high science" text from the *Journal des Savants*.

### Corpus description

The *Bibliothèque utile* was founded by the republican parliamentarian Henri-Charles Leneveux<sup>2</sup> in 1859. Twenty volumes were published in 1859 and 1860, twelve in 1860 and 1861, and it was published thereafter at intervals until at least 1903.

The *Bibliothèque utile* was "consacré à la vulgarisation des connaissances les plus indispensables à l'homme et au citoyen" ("devoted to the popularisation of the knowledge that is the most indispensable for man and the citizen"), printed on the back cover of each volume, and a 1948 governmental study entitled "La Lecture Publique en France" states (Secrétariat général du gouvernement, 4):

des collections à bon marché comme celle de "La Bibliothèque utile", petits in-18 (sic<sup>3</sup>) de 192 pages à 0 fr. 50 imprimés par le typographe Leneveux ... remportent auprès des classes laborieuses un vif succès.

inexpensive collections like "The Bibliothèque utile", small in-18, 192 pages costing 50 cents, printed by the typographer Leneveux ... are highly successful with the working classes.

---

<sup>2</sup> Henri-Charles Leneveux also co-authored the *Bibliothèque utile* volume n° 23 "L'instruction en France" in 1861.

<sup>3</sup> In fact, their format was a handy-sized in-16, measuring only 15 cm by 10 cm.

It certainly acquired a good reputation as, in 1862, Georges Sand wrote, in correspondence with her niece (Sand, letter DXIX):

la *Bibliothèque utile* de Leneveu (sic), ... est un excellent recueil, très répandu et très goûté.

Leneveu's *Bibliothèque utile* ... is an excellent collection, very widespread and much appreciated.

As for the other contributors, there were many different authors, one author per issue. For example, volume n° 62 entitled “The sciences” originally written in English by Thomas Huxley, was translated for the *Bibliothèque utile* by a certain Henry Gravez.

The first volume of the *Bibliothèque utile* in 1859, entitled “Introduction à l'étude des sciences physiques” (“Introduction to the study of physical science”) was written by Joseph Morand, a teacher. He had already written a long “Histoire philosophique des sciences et de la civilisation” (“Philosophical history of science and civilisation”) in 1838 (614 pages). He was also to go on to write other works, like “Sciences physico-chimiques” (“Physical and chemical sciences”), in 1861. In addition, he later had his *Bibliothèque utile* text re-published, with the same title, by Dubuisson, in 1862 (188 pages).

In contrast, the *Journal des Savants* was founded in 1665. Its standing is similar to that of *The Philosophical Transactions of the Royal Society of London* (originally *Philosophical Transactions*), first published in 1665, which became the official publication of the Royal Society only in 1752. Like the *Philosophical Transactions*, the *Journal des Savants* was written by researchers, scientists and other educated gentlemen, to be read by their peers.

The renowned French physicist, mathematician and astronomer, Jean-Baptiste Biot<sup>4</sup> wrote 5 chapters of approximately 20 pages each, in 5 consecutive issues of the *Journal des Savants* from October 1846 to February 1847 on the topic “Sur la planète nouvellement découverte par M. Le Verrier, comme conséquence de la théorie de l'attraction” (“On the planet newly-discovered by Mr Le Verrier, as a consequence of the theory of attraction”). Biot describes the work of Kepler, and the later work of Newton which would lead to the discovery of this planet: Neptune.

The *Bibliothèque utile* is divided into short sections, so I chose one section from the *Bibliothèque utile* on a topic that was also available in the

---

<sup>4</sup> Jean-Baptiste Biot was a Professor of mathematical physics, first at the Collège de France and later at the Faculty of Science, Paris.

*Journal des Savants*, namely Kepler's laws of planetary motion and, within that subject, the corresponding pages: 10 pages in the *Bibliothèque utile* and 9 in the *Journal des Savants* (around 3000 words in both sections).

## Grammatical metaphor

Grammatical metaphor is an important notion in Systemic Functional Linguistics, and is especially useful when examining and analysing scientific texts. For readers unfamiliar with this concept, I cite David Banks (Banks, 2017, 77):

...in traditional metaphor, a word of one class is replaced by another word of the same class, but with a different meaning. In grammatical metaphor, it is the other way round. We replace one word by a word of a different class, but that has the same basic meaning.

So, for example, instead of using verbs to express processes, which SFL calls the congruent form, nominalisation (the non-congruent form) can be used. Processes are often presented as verbs in the first instance, then nominalised in the subsequent text. In my corpus, only ideational grammatical metaphors are considered, since this is the type encountered most commonly in scientific discourse, and are hereafter referred to as “grammatical metaphors”. I also exclude any grammatical metaphors that have become the unmarked choice for expressing the process in the language, like “à mesure que” (“as”) present in my corpus.

An example of the ideational type of grammatical metaphor can be seen below (BU, 53), as Morand describes the planets (italics in the original text; my highlighting in bold; my translations and glosses):

a) Nous appelons *principales* toutes celles qui, comme la terre, **tournent directement autour du soleil**, et *secondaires* celle qui, comme la lune, **notre satellite, exécutent leur mouvement de translation** autour d'une planète principale, et leur **mouvement de rotation** par rapport au soleil.

We call *principal* those (planets) that, like the earth, **turn directly around the sun**, and *secondary* those that, like the moon, **our satellite, perform their movement of translation** around a principal planet, and their **movement of rotation** in relation to the sun.



In example (a), the verbal expression “**toument ... autour de**” is re-used in nominalised form “**mouvement de translation**” in a construction using the “empty” verb “**exécutent**”.

Similarly, the process involved in rotation is also presented first as a verb, then nominalised (BU, 54):

b) **Toutes les planètes qu'on a été à même d'observer tournent sur leur axe, et les idées générales qui résultent de l'ensemble des faits constatés ne permettent d'admettre, à l'égard du mouvement de rotation, aucune exception probable dans le système planétaire.**

All the planets that we have been able to observe turn on their axis, and the general ideas that result from all the facts noted do not allow us to assume, concerning the movement of rotation, any probable exception in the planetary system.

where “**tournent sur leur axe**” becomes “**mouvement de rotation**” in the following proposition.

However, as a counter example, in example (c), describing the transit of Venus, the nominalised “**passage**” precedes the process “**passera**” in the text (BU, 54):#

c) ... **lors de ses passages sur le disque solaire en 1761 et 1769. On croit qu'elle passera de nouveau entre nous et le soleil en 1874 et 1882.**

... during its passing across the sun's disc in 1761 and 1769. We believe that it will pass again between us and the sun in 1874 and 1882.

In the *Journal des Savants* (JdS) the following example of grammatical metaphor also shows the process initially encoded congruently as a verb, then re-introduced as a nominalisation in the next sentence (JdS, 587):

d) **Conformément à l'hypothèse jusqu'alors universellement admise, que les astres doivent se mouvoir dans des cercles, avec des vitesses constantes, il (Kepler) détermine le point central du cercle de Mars, et prouve que toutes les longitudes constatées par l'observation se trouvent ainsi représentées aussi exactement qu'elles peuvent l'être ; puis il montre que les latitudes calculées, qui y correspondent, discordent avec les latitudes observées. Donc l'hypothèse de la circularité est démontrée fausse.**

In accordance with the hypothesis up until now universally accepted, that the stars must move in circles, with constant speeds, he (Kepler) determines

the central point of the circle of Mars, and proves that all the longitudes noted by observation are found thus represented as exactly as they can be; then he shows that the latitudes calculated, which correspond to them, are discordant with the latitudes observed. Therefore the hypothesis of circularity is shown to be false.

The total number of all grammatical metaphors in the two texts are as follows. *Bibliothèque utile*: 120 occurrences (4% of the total number of words) and the *Journal des Savants*, 196 occurrences (6% of the total number of words), which means that the *Journal des Savants* text has a higher number of nominalisations.

A further main difference between the *Bibliothèque utile* and the *Journal des Savants* is the presence, or not, of long nominal groups with iterated phrase-and-group qualifiers, sometimes referred to as “embedding” in SFL. In the *Bibliothèque utile*, the most common type of complex nominal group is composed of two lexical items, like “la somme des distances” (the sum of the distances”), “les durées de leurs évolutions” (“the durations of their evolutions”), “la méthode d'exposition” (“the method of exposition”), and the only complex nominal group in the corpus (BU, 66) is the following:

e) autres erreurs [relatives à la détermination [des orbés et de leurs mouvements]]<sup>5</sup>.

other errors relative to the determination of the planets and of their movements.

In contrast, the *Journal des Savants* contains many longer, more complex nominal groups, like this example with 4 embedded phrases (JdS, 591):

f) les variations [d'intensité [de la pesanteur [en diverses parties [de leurs surfaces]]]]

the variation in intensity of the weight in various parts of their surfaces

At this point we can examine the lexical density of this corpus. Halliday measures lexical density by dividing the number of lexical items per sentence by the number of ranking clauses. Unplanned spoken language is

---

<sup>5</sup> Note that the markers [ ] indicate an embedded phrase. So here, the final ]] indicates that there are two embedded phrases.

“more complex in terms of grammatical intricacy” (Halliday 1994, 351); the lexical items are distributed over a larger number of grammatically related clauses so lexical density is typically 1.5 or 2 (Halliday 1989, 80). But in written language, lexical items are concentrated more frequently in nominal groups: “the nominal group is the primary resource used by the grammar for packing in lexical items at high density” (Halliday 1994, 351). In written language lexical density for English is between 3 and 6, rising to 10 or more for highly planned contemporary scientific writing (Ormerod 2000, 76).

In the *Bibliothèque utile*, with its shorter clauses and fewer ranking clauses, lexical density is typically 5 or 6. However, even though the *Journal des Savants* text contains many longer clauses with 10 to 16 (and occasionally more) lexical items in total, the lexical density is not higher. There are even clauses in the text, like the example below, that might appear difficult to comprehend for the general reader. The following contains 17 lexical items over 3 clauses, so has a lexical density of just under 6 (JdS, 590):

g) Il ne peut en résulter que des effets mécaniques, d'un ordre de petitesse relatif, qui permet d'en remettre l'appréciation à des approximations ultérieures, pour les ajouter, comme conséquences de la même force, aux phénomènes d'ensemble, quand ils seront calculés.

From this, there can result only mechanical effects, of a relative order of smallness, which allows (us to) postpone appreciation of them to subsequent approximations, in order to add them, as consequences of the same force, (on) to the overall phenomena, when they are calculated.

Contrary to my expectations, lexical density does not therefore seem to be a relevant criterion for assessing whether the “low science” text, the *Bibliothèque utile*, is more complex than the “high science” text, the *Journal des Savants*.

## Grammatical metaphor and Theme

In SFL, Theme is the term used to denote the beginning of the message and is found in initial position in the sentence. Halliday (1994, 38) defines it thus:

The Theme is one element in a particular structure which, taken as a whole, organizes the clause as a message; this is the configuration Theme + Rheme. A message consists of a Theme combined with a Rheme.

Fries (1995, 349) further elaborates:

The Theme of a clause is the initial constituent of that clause and provides a framework for the interpretation of the message expressed by the clause.

The Rheme is the rest of the sentence. In an unmarked context, Old information is generally presented at the beginning of a sentence, linking back to previous information, and New information is then introduced after the Theme.

Grammatical metaphor has an important role in writing as it can encapsulate information from the previous text, and present it in Theme position as Old information, enabling New information to be introduced naturally in the Rheme.

The extract chosen from the *Bibliothèque utile* does not contain any examples of grammatical metaphor being used in Theme position to present New information in the Rheme. However, in the *Journal des Savants* (JdS, 587), example (d) above shows how the nominalisation “la circularité” appears in the Theme of the following sentence.

h) **Donc l'hypothèse de la circularité est démontrée fausse.**

The next example from the *Journal des Savants* shows how the grammatical metaphor “écarts” is built up over several clauses and introduced by the deictic “De tels” (JdS, 592) (author's italics):

i) Mais, la gravitation étant universelle, ces conditions simples ne sauraient exister, et chaque planète doit être incessamment **écartée de son ellipse solaire propre** par les attractions que tous les autres corps planétaires exercent sur elle, selon leurs masses, leurs positions et leurs distances. **De tels écarts** se constatent en effet, et on les nomme les *perturbations planétaires*.

But, gravitation being universal, these simple conditions cannot exist, and each planet must be incessantly **diverted from its own solar ellipse** by the attractions that all the other planetary bodies exert on it, according to their masses, their positions and their distances. **Such diversions** are indeed noted, and we call them *planetary perturbations*.

In the above example, we also have the nominal group “ces conditions simples”, which is an instance of what Gill Frances (1994) calls retrospective labelling. We examine this in the following section.

## Retrospective labels

Unlike grammatical metaphors which package previous information into a nominalised process, retrospective labels package previous information, sometimes lengthy descriptions or explanations, into shorter anaphoric nominal groups signalled by a deictic (Francis, 1994, 85) as in example (i) “ces conditions simples”:

The head nouns of retrospective labels are almost always preceded by a specific deictic like *the, this, that* or *such*, and may have other modifiers and qualifiers too.

As we saw earlier, example (b) (BU, 54) contains grammatical metaphor, but also retrospective labelling (in bold below):

j) Toutes les planètes qu'on a été à même d'observer tournent sur leur axe, et les idées générales qui résultent de l'ensemble des faits constatés ne permettent d'admettre, à l'égard du mouvement de rotation, aucune exception probable dans le système planétaire.

However, the *Journal des Savants* contains a greater number of examples of retrospective labelling (JdS, 592):

k) L'individualité des mouvements s'observe surtout, presque complète, dans les trois systèmes partiels que je viens de désigner. En s'appuyant sur ce fait, on peut calculer les rapports des masses de leurs planètes centrales à la masse du soleil.

The individuality of the movements can be observed especially, almost complete, in the three partial systems that I have just indicated. Relying on this fact, we can calculate the ratios of the masses of their central planets to the mass of the sun.

and here (JdS, 593):

l) Alors le rapport des forces exercées ainsi à une même distance vous donnera le rapport de la masse de Jupiter à la somme des masses de Jupiter et du soleil, d'où vous déduirez le rapport de ces deux dernières individuellement. Un calcul pareil vous donnera le rapport analogue pour Saturne et pour Uranus.

So the ratio of the forces exerted thus at a same distance will give you the ratio of the mass of Jupiter to the sum of the masses of Jupiter and (of) the sun, from whence you will deduce the ratio of these latter two individually.

A **similar calculation** will give you the **analogous ratio** for Saturn and for Uranus.

where “ce fait” encapsulates all the information from the previous sentence, and “un calcul pareil” avoids repeating the mathematical process elaborated in the previous sentence.

The lexical item “fait” (“fact”) in this nominal group is one of the labels in Francis's list of the most common general labels found in her corpus from *The Times* newspaper (Francis 1994, 89). In the *Bibliothèque utile* and *Journal des Savants* corpus, the most common retrospective labels (singular and/or plural) comprise mainly science-oriented items:

analyse, calcul, circonstance, conjecture, conséquence, construction, déduction, distribution, fait, hypothèse, méthode, phénomène, présomption, procédé, proportion, résultat.

analysis, calculation, circumstance, conjecture, consequence, construction, deduction, distribution, fact, hypothesis, method, phenomenon, presumption, procedure, proportion, result.

## Interpersonal interaction

In twentieth century scientific discourse, one does not expect to find much, if any, explicit interpersonal interaction between the author and his/her reader. However, in the *Bibliothèque utile*, there are several instances where Morand addresses his readers, to encourage them to understand what he means, to assure the ladies in their salons and the men in their workplaces that science is easy, in an attempt to make science accessible to all (BU, 59):

m) On peut donner, même à des lecteurs peu éclairés, une idée sommaire des trois grandes lois astronomiques formulées par Kepler.

We can give, even to unenlightened readers, a summary idea of the three great laws of astronomy formulated by Kepler.

After a very laborious description of how to draw an ellipse, which also appears above the text as a diagram, labelled Fig. 3, he adds (BU, 60):

n) Cette courbe est une ellipse. Sa construction n'est sûrement pas difficile à comprendre, et il ne faut qu'une attention médiocre pour bien saisir ce que nous allons dire.



This curve is an *ellipsis*. Its construction is surely not difficult to understand, and it requires only mediocre attention to really understand what we are going to say.

Morand also presents unsubstantiated scientific claims that would not be acceptable in a scientific article today, even in a popularised review. The following are 3 examples taken from this small corpus, but there are many others in this book<sup>6</sup>. The first assumes that Venus is inhabited (BU, 55):

○) C'est par la vitesse et la direction de ces tâches qu'on a trouvé que la durée de la rotation de Venus est un peu moindre que celle de la terre et que les habitants de cette planète ont des jours de moindre durée que les nôtres.

It is through the speed and the direction of these spots that it has been found that the length and the rotation of Venus is slightly less than that of earth's and that the inhabitants of this planet have days of a length shorter than ours.

The second example describes Jupiter (BU, 56):

p) Il n'est pas croyable qu'une planète mille fois plus grosse que notre globe, ... ne soit pas habitée, et ... son règne organique doit être incomparablement plus étendu, plus riche, plus varié, plus parfait que le nôtre.

It is not credible that a planet a thousand times larger than our globe, ... should not be inhabited, and ... its organic kingdom must be incomparably more extensive, richer, more varied, more perfect than ours.

and when Morand writes about the ring (sic) surrounding Saturn (BU, 57):

q) Qu'est-ce que cet anneau ? De quelle manière est-il formé ? L'homme n'en saura jamais rien.

What is this ring? How is it formed? Mankind will never know.

In contrast, in the *Journal des Savants* there are no unsubstantiated scientific claims. There is no need to explain how to draw an ellipsis. Biot gives no explanations of a complex mathematical nature, instead leaving it to the

---

<sup>6</sup> (BU, 41): “La lune est habitée, nous n'en doutons nullement et, dans notre opinion, ce n'est pas seulement la lune qui est habitée, mais aussi toutes les autres planètes, quel qu'en soit le nombre.” The moon is inhabited, we have no doubt about it and, in our opinion, it is not only the moon that is inhabited, but also all the other planets, whatever their number.

scientist reading the article to comprehend the only table that appears in the text, saying (JdS 594):

r) Le tableau suivant offre l'ensemble des évaluations ainsi obtenues, ... On comprendra ainsi par les nombres, mieux que les paroles, ...

The following table offers all the evaluations thus obtained, ... So we/you will understand from the numbers, better than words, ...

The only remarks of an interpersonal nature occur in one paragraph in the chosen extract where Biot imagines Kepler's state of mind when he made his discoveries (JdS, 588):

s) Kepler, pauvre jusqu'à la misère, dut ressentir une grande joie, sans doute, quand il se vit possesseur de ces admirables découvertes, qui rassemblaient tous les résultats astronomiques dans leur énoncé.

Kepler, poor to the point of destitution, must have felt great joy, without doubt, when he found himself the possessor of these admirable discoveries, which brought together all the astronomical results in their wording/formula.

### By way of conclusion

My hypothesis when undertaking this study was that there would be a marked contrast between the popularised science discourse, “low science” text, in the *Bibliothèque utile*, compared with the academic scientific discourse, “high science” text, in the *Journal des Savants*.

To a certain extent, that has proved true. Nominalisations and grammatical metaphor are common in the two texts, but much more frequent in the *Journal des Savants*. Complex nominal groups are not very common in either text, but more common in the *Journal des Savants*. Likewise, retrospective labels are more frequent in the *Journal des Savants*. However, lexical density is similar in both texts

The greatest contrast is to be found not only in the interpersonal aspects of the texts, omnipresent in the *Bibliothèque utile*, absent from the *Journal des Savants*, but also in the scientific content. The *Bibliothèque utile* makes unsubstantiated scientific claims, again absent from the *Journal des Savants*, which would be unacceptable in either popular or academic publications today.

## Corpus

- Biot, Jean-Baptiste. 1847. "Sur la planète nouvellement découverte par M. Le Verrier, comme conséquence de la théorie de l'attraction". *Journal des Savants*. Paris: Imprimerie Nationale.
- Morand, J. 1859. "Introduction à l'étude des sciences physiques". *Bibliothèque utile* n°1. Paris: Librairie Pagnerre.

## References

- Banks, David. 2017. *A Systemic Functional Grammar of French*. London: Routledge.
- Francis, Gill. 1994. "Labelling Discourse: an Aspect of Nominal-Group Lexical Cohesion". In *Advances in Written Text Analysis*, edited by Malcolm Coulthard, 83-101. London: Routledge.
- Fries, Peter H. 1995. "Themes, Methods of Development, and Texts". In *On Subject and Theme: A Discourse Functional Perspective*, edited by Ruqaiya Hasan and Peter H. Fries, 317-359. Amsterdam: John Benjamins.
- Halliday, M. A. K. 1989. *Spoken and Written Language*. Oxford: Oxford University Press.
- Halliday, M. A. K. 1994. *An Introduction to Functional Grammar*. 2<sup>nd</sup> edition. London: Edward Arnold.
- Halliday, M. A. K. & C. M. I. M. Matthiessen. 2014. *An Introduction to Functional Grammar*. 4<sup>th</sup> edition. London: Arnold.
- Halliday, M. A. K. and Martin, J. R. 1993. *Writing Science: Literacy and Discursive Power*. London: The Falmer Press.
- Marchat J.-F. 2008. "Bucheux et Le Play : approches éducatives de deux écoles d'économie sociale catholiques". *Le Télémaque* n°33: 45-60. Caën: Presses Universitaires de Caën.
- Ormerod, Janet. 2000. "La rédaction de l'article scientifique par des anglophones et par des francophones: le groupe nominal". Unpublished doctoral thesis, Université de Bretagne Occidentale.
- Ormerod, Janet. 2009. "La métaphore grammaticale idéationnelle : étude préliminaire de son emploi en français". In *La Linguistique Systémique Fonctionnelle et la Langue Française*, edited by David Banks, Simon Eason and Janet Ormerod, 157-175. Paris: L'Harmattan.

### Online references

Sand, George. 1862. "Letter DXIX to Mademoiselle Nancy Fleury", *Correspondance 1812-1876*. Paris: Calmann-Lévy (1883). Accessed 28/01/2019:

[https://fr.wikisource.org/w/index.php?title=Correspondance\\_1812-1876,\\_4/Texte\\_entier&printable=yes](https://fr.wikisource.org/w/index.php?title=Correspondance_1812-1876,_4/Texte_entier&printable=yes)

Secrétariat général du gouvernement. 1948. "La lecture publique en France". Accessed 28/01/2019:

<https://www.enssib.fr/bibliotheque-numerique/documents/48809-notes-documentaires-et-etudes-n-918-serie-francaise-cxcii-la-lecture-publique-en-france.pdf>.

CHAPTER SIX

METADISOURSE AND STANCE-TAKING  
IN PREFACES:  
A DIACHRONIC ANALYSIS

MOHAMED SAKI

**Introduction**

The aim of this article is to explore metadiscourse, the language of evaluation and stance-taking in six prefaces to novels written in the 18<sup>th</sup>, 19<sup>th</sup> and 20<sup>th</sup> centuries. It is my contention that metadiscourse is an interpersonal practice through which writers take a stance towards what they say as well as towards their audiences. Therefore, I will study the interpersonal dimension of the metadiscourse used in these prefaces so as to shed light on how their writers position themselves both towards their central text and their potential readers and how they engage with them. In addition, I will analyse the crucial role played by metadiscourse in helping the authors to project themselves on to their unfolding texts and to manage various communicative intentions. By analysing metadiscourse and stance-taking in these prefaces, I will also investigate how the preface writers use the generic characteristics as well as the illocutionary force of the preface as a genre, and how they set an interpretive grid and assign positions from which the would-be readers are invited or encouraged to appraise their novels. Finally, the aim of this article is to see if the preface, as a genre, has evolved and to what extent, and to determine what differences and similarities, if any, in the uses of metadiscourse can be distinguished, even if the corpus is limited. Before analysing the six prefaces, I will first start by introducing the two concepts I will be working with: the *preface* and *metadiscourse* so as to present the theoretical assumptions that inform my analysis.

## 1. The preface and metadiscourse

### 1.1 The preface

The preface is not a transparent and straightforward appendix to a book; it is rather a communicative event of a fundamentally metadiscursive nature whose aim is twofold: promotional and justificatory (Genette, 1978, Saki, 2013). The preface seeks to initiate to the knowledge of the text by providing information that may help better understand the central text and, sometimes it may address, either directly or not, a specific audience. It also aims to explain, overtly or explicitly, why the text it introduces is worth reading and how it should be read as well. Therefore, the preface, as a subgenre, presupposes that the preface writer takes a stance towards the “central” text and towards the targeted readers.

Alongside the title, the epigraphs, the table of contents, etc., the preface belongs to what G. Genette calls the *peritext*, i.e. those textual elements which share the same printable space as the “central” text and from which they are not “physically” dissociated. G. Genette qualifies the preface as a *transitional space* and a *transactional space*; it is *transitional* because it is a threshold, located between the text and the “off-text”; i.e. what is outside the text and not related to it; it is *transactional* because it is a contact zone where authors negotiate a specific relationship with a more or less clearly delineated community of (ideal) readers (Genette 1987: 374).

### 1.2. Metadiscourse

Metadiscourse is a complex phenomenon which may be broadly defined as the linguistic material of the text that is separated from its propositional content; it is text about text, or discourse about discourse. According to Hyland, metadiscourse consists of those linguistic tokens in a text “which explicitly refer to the organization of the discourse or the writer’s stance towards either its content or the reader” (Hyland, 2005: 438). In other words, metadiscourse includes the non-propositional linguistic material that refers to the writers themselves, their audiences, or their evolving texts, but not to the ideas their texts convey. In any case, metadiscourse signals “the author’s intrusion into the discourse, either explicitly or non-explicitly, to direct rather than inform the readers” (Crismore, 1989, 4).

Ädel (2006) makes a distinction between an “integrative approach” and a “non-integrative approach”. The former sees metadiscourse as “the means whereby the writer’s presence in the discourse is made explicit, whether by displaying attitude towards or commenting on the text or by showing how



the text is organized" (Ådel, *ibid.* 168). The latter considers that metadiscourse "primarily investigates aspects of text organization, while largely excluding interpersonal elements" (*ibid.* 175).

For Makkonen-Craig, metadiscourse is, first, text-oriented, because it refers to the evolving text and, second, participant-oriented, because it refers to the current writer and the imagined current reader. The current writer is the organizer of the text who interacts with the imagined reader in the world of discourse. Consequently, the metadiscourse material of a text fulfils two rhetorical functions: guidance and interaction (2011, 686).

As for Vandekopple (1997) metadiscourse does not expand referential material, it primarily helps readers connect, organize, interpret, evaluate, and develop attitudes toward the unfolding text they are reading. In sum, metadiscourse seeks to make readers understand the text and to make the writer's intention more explicit.

What stands out from the different definitions of metadiscourse is that it is fundamentally both self-reflexive and interpersonal. Metadiscourse indicates how writers take a stance towards their text – by organising, classifying, glossing, or evaluating it and how they negotiate an intersubjective relationship with their readers (Hyland and Tse, 2004).

It is not always easy to pin down the metadiscursive material in a text to a limited set of linguistic elements; this material may draw from an open-ended set of language items that can also perform non-metadiscourse roles. In fact, actual metadiscursive markers can only be recognized through empirical analysis of the text.

This difficulty largely accounts for the different systems that have been elaborated to classify metadiscursive devices and their uses. Vandekopple classifies metadiscourse markers into two categories: *textual* – text connective, code glosses, validity markers, narrators – and *interpersonal* illocution markers, attitude markers, commentaries, (1985, 86-87). But this classification has been criticized and several shortcomings and inconsistencies have been pointed out, especially by Hyland, who elaborated an alternative system of classification (Hyland, 2009).

Hyland considers that the textual-interpersonal division of metadiscourse is ineffective since metadiscourse refers chiefly to elements of the text that signal writer-reader interaction. In addition, he maintains that even what may be viewed as purely textual items can sometimes fulfil an interpersonal function or convey a propositional meaning (Hyland and Tse, 2004: 167).

Hyland's system of classification of metadiscourse is based on the distinction between the *interactive* and the *interactional* dimensions of metadiscourse. The *Interactive resources* "allow the writer to manage the information flow to explicitly establish his or her preferred interpretations"

while the *interactional resources* “focus on the participants of the interaction and seek to display the writer’s persona and a tenor” so as to establish a suitable relationship to their text and their readers (Hyland, 2009, 129).

This system of classification is presented in the table below:

Table 1 metadiscursive resources

<b>Interactional resources</b>	
Hedges	- mark the writer’s reluctance to present propositional information categorically, downplay commitment, open dialogue;
Boosters	- express certainty, emphasise the force of propositions, close dialogue.
Attitude markers	- express the writer’s appraisal of propositional information;
Self-mentions	- signal the author’s presence in terms of first person pronouns and possessives, explicit reference to the author;
Engagement markers	- address readers, explicitly or implicitly, by focusing their attention or by including them as participants, second person pronouns, directives, question forms and asides, etc.
<b>Interactive resources</b>	
Transitionals	- indicate relations between main clauses;
Frame markers	- indicate text boundaries, label text stages, announce discourse goals;
Endophoric markers	- make additional material salient, refer to other parts of the text;
Evidentials	- indicate information sources which originate outside the current text;
Code glosses	- indicate the restatement of ideational information.

My analysis of the metadiscourse material in the three prefaces will draw mainly on Hyland’s classification whose basic assumption – that metadiscourse is basically interpersonal – is one of the premises of this article.

## 2. The metadiscursive devices in the three prefaces

My corpus is composed of six prefaces: two written by Henry Fielding, to *Joseph Andrews* (1742) and *Tom Jones* (1749), two by Nathaniel Hawthorne, to *The Scarlet Letter* (1850) and *The House of Seven Gables* (1851), and two by Vladimir Nabokov, to *Bend Sinister* (1933) and *The Defense* (1964). These prefaces are various lengths, ranging from one to seven pages. This choice is highly random since I have not chosen to strike a balance between American authors and British authors; this limited corpus meets, however, two criteria: the prefaces are authorial – i.e. written by the novelists themselves – and they were written over a time span of almost two hundred years. The main purpose of this study is concerned exclusively with the permanent and changing generic features of the preface as a subgenre.

### 2.1. The interactional resources

The aim behind analysing the interactional resources of these prefaces is to grasp how the authors construct their own textual persona, or ethos, how they establish a particular relationship to their readers and how they position themselves with regard to the novels they introduce. The interactional resources that predominate in the six prefaces are self-mentions, attitude markers and engagement markers. There are almost no hedges in these texts and, consequently, most statements are endowed with a strong commitment and involvement of the textual persona of the prefaces.

#### *i- Self-mentions*

The preface is a contact zone where writers address their potential readers, in their own name and not behind any fictional and narrative mask; the authorial presence can be perceived, mainly, through the use or absence of self-mentions. It is worth analysing what linguistic forms these metadiscursive items take. Self-mentions in my corpus take two forms: a straightforward one and an indirect one. Fielding and Nabokov refer directly to themselves by using the first person singular pronoun *I* and its affiliates, *my* and *me*; I have found 11 occurrences in *Joseph Andrews* and only 6 in *Tom Jones* but 20 occurrences in *Bend Sinister* and 13 in *The Defense*:

1. *Bend Sinister* was the first novel *I* wrote in America, and that was half a dozen years after she and *I* had adopted each other. (*Bend Sinister*, p. 5)
2. *I* might observe that our Ben Jonson, who of all men understood the Ridiculous the best, hath chiefly used the hypocritical affectation. (*Joseph Andrews*, p. 5)

The uses of the first person singular pronouns are associated with personal and aesthetic considerations that both authors weave through their prefaces to provide their readers with the appropriate background to their texts, to expose their own conception of art in general and of the art of the novel in particular. One of the main differences between Fielding's and Nabokov's prefaces is that in the latter, the author tells his would-be readers why they should not mistake him for any of his characters. Both authors present themselves as the primary sources of crucial information readers should have in order to appraise their novels correctly.

The second form of self-mentions is indirect and it consists of using *the author*, and the third person pronoun *he* and its affiliates *his* and *him*. I have found 15 occurrences in Fielding's prefaces, 26 in Hawthorne's and 6 in Nabokov's:

3. As it is possible the mere English reader may have a different idea of romance with the **author** of these little volumes; (*Joseph Andrews*, 6)
4. An **author** ought to consider himself, not as a gentleman who gives a private or eleemosynary treat, but rather as one who keeps a public ordinary, at which all persons are welcome for their money (*Tom Jones*, 4)
5. Much to the **author's** surprise, and (if he may say so without additional offence) considerably to his amusement, he finds that his sketch of official life, introductory to *The Scarlet Letter*, has created an unprecedented excitement in the respectable community immediately around him. (*The Scarlet Letter*, 3)
6. In the present work the author has proposed to himself -but with what success, fortunately, it is not for him to judge - to keep undeviatingly within his immunities. (*The House of Seven Gables*, 7)
7. It may be asked if it is really worth an **author's** while to devise and distribute these delicate markers whose very nature requires that some that they not be too conspicuous. (*Bend Sinister*, 10)

Finally, another device of self-mention consists of using the royal 'we' and its associates *our* and *us*; this self-mention device is found only once in Nabokov's prefaces and 10 times in Fielding's; Hawthorne does not use it at all:

8. But tho' **we** have sometimes admitted this in our diction, **we** have carefully excluded it from *our* sentiments and characters (*Joseph Andrews*, 4)
9. Is there any judgment on my part carried out, any sentence pronounced, any satisfaction given to the moral sense? If imbeciles and brutes can punish other imbeciles and brutes, and if crime still retain an objective

meaning in the meaningless world of Paoluk.... We affirm that crime is punished at the end of the book... (*Bend Sinister*, 7-8)

In all, there are 107 self-mentions in our corpus as we can see in table 2 below:

Table 2: self-mentions

Self-mention	Nabokov	Hawthorne	Fielding	Total
I, my	33	∅	17	50
Us, our	10	∅	∅	10
The author, he, him	6	26	15	47
Total	49	26	32	107

The 107 occurrences of self-mentions in these prefaces display differences; the three authors do not necessarily use the same metadiscursive devices of self-mentions and therefore there are some important idiosyncrasies. Still, in spite of these differences, these metadiscursive choices contribute to stage the authors' interventions in the prefaces and to construct a textual persona. They refer both to the authors themselves, *qua* empirical beings, and to their doubles: their own literary critics. The duality of the self-mentions signals, by the same token, the type of relationship the three authors aim to establish with their readers, a relationship characterized by intimacy and distance. They insist on intimacy and proximity when they use the first person pronouns and they enact distance and remoteness when they use the third person pronouns and impersonal self-references.

*ii- Attitude markers*

Attitude markers express the writers' appraisals of the propositional information of their text and they help reveal the stance being taken towards it. The analysis of the attitude markers of the prefaces of my corpus brings to the fore how each of the three authors encodes his opinion about his novels and how he presents them to his readers. It also shows how the double status of the textual personas they construct in their prefaces – authors and critics at the same time – enables them to appraise their novels and, obviously enough, to do so positively:

10. But perhaps it may be objected to me, that I have against my own rules introduced vices, and of a very black kind into this work. To this I shall answer: First, that it is very difficult to pursue a series of human actions and keep clear from them. Secondly, that the vices to be found here, are rather the accidental consequences of some human frailty, or foible,

than causes habitually existing in the mind. Thirdly, that they are *never* set forth as the objects of ridicule, but detestation. Fourthly, that they are *never* the principal figure at that time on the scene; lastly, they never produce the intended evil. (*Joseph Andrews*, 5)

11. But it appears to him, that the only remarkable features of the sketch are its frank and genuine good-humor, and the general accuracy with which he has conveyed his sincere impressions of the characters therein described. (*The Scarlet Letter*, 3)
12. The book teems with stylistic distortions, such as puns crossed with anagrams, suggestive neologisms, ...parodies of narrative clichés... spoonerisms...; and of course the hybridization of tongues. (*Bend Sinister*, 8-9)

Attitude markers in these prefaces, such as the ones in bold characters in the examples above, highlight a fundamental function of the preface as a genre: it is the semantic locus where the textual persona, the writer's double, reinforces its status as the authorized critic and insider pundit, a "textual gate-keeper" who can spell out overtly the original intentions of the writer. Nabokov states it bluntly when he says that: "in the long run, however, it is only the author's satisfaction that counts." (*Bend Sinister*, 10)

### iii) Engagement Markers

Engagement markers seek to focus the attention of the potential readers by addressing them, directly or indirectly, thus constructing them as discourse participants. These markers can be pronouns, (you or inclusive 'we'), interjected phrases, ("you may notice", "it is worth noticing", etc.), informational or rhetorical questions, directives, indirect addresses (the reader), etc. Engagement markers are obviously of high importance when we investigate the interpersonal dimension of the prefaces and the reading positions the writers embed in them.

The analysis of the engagement markers in the six prefaces shows that Fielding and Hawthorne interact in a very similar way with their would-be readers, while there is a marked difference with Nabokov. Fielding, for instance, addresses, in both his prefaces, a general, metonymic reader:

13. As it is possible the mere English reader may have a different idea of romance with the author of these little volumes; and may consequently expect a kind of entertainment, not to be found, nor which was even intended, in the following pages. (*Joseph Andrews*)
14. An author ought to consider himself, not as a gentleman who gives a private or eleemosynary treat, but rather as one who keeps a public ordinary, at which all persons are welcome for their money. (*Tom Jones*, 5)



Hawthorne addresses only those readers who reacted violently to the first edition of his novel in the preface to *The Scarlet Letter* (15) and the general public in the preface to *The House of Seven Gables* (16)

15. As to enmity, or ill-feeling of any kind, personal or political, he utterly disclaims such motives. (*The Scarlet Letter*, 3)
16. The reader may perhaps choose to assign an actual locality to the imaginary events of this narrative. If permitted by the historical connection, -which, though slight, was essential to his plan, -the author would very willingly have avoided anything of this nature. (*The House of Seven Gables*, 6)

In his two prefaces, Nabokov systematically makes a distinction between two categories of readers and establishes with them two binary interpersonal relations. On the one hand, a group of general, unidentified readers, and, on the other, a group of readers he wards off and whom he, quite often, turns into derision. He establishes with the first group of readers, whom he instructs on how to read his novels, an impersonal and distant interpersonal relation; he refers to them vaguely and generically:

17. The author trusts that wide readers will refrain from avidly flipping through his autobiography *Speak, Memory* in quest of duplicate items or kindred scenery. The fun of *Glory* is elsewhere. It is to be found in the echoing and linking of minor events... the reader can only guess at, even after he has raced through the last seven chapters... (*The Defense*, 11)
18. When in Chapter Three, Ember alludes to three best-selling novels, the alert commuter cannot fail to notice that the titles of three of them form, roughly, the lavatorial injunction not to Flush the Toilet when the Train Passes through Towns and Villages... (*Bend Sinister*, 9-10)

With the second subcategory of readers, Nabokov openly imposes an asymmetrical interpersonal relation with his would-be readers, with whom he refuses to establish collegiality: the different engagement markers he resorts to contribute to fencing his novels and warding off “undesirable readings”:

19. In the Prefaces I have been writing of late for the English-language editions of my Russian novels... I have made it clear to address a few words of encouragements to the Viennese delegation... the little Freudian who mistakes a Pixlok set for the key to a novel will no doubt continue to identify my characters with his comic-book notions of my parents, sweethearts and serial selves. For the benefit of such sleuths, I may as well confess that I gave Luzhin my French governess, my pocket chess...” (*The Defense*, 10-11)

20. All my books should be stamped **Freudians**, Keep **Out**. (*Bend Sinister*, 10)

Engagement markers referring to the second group of readers are more frequent in his prefaces. They target two distinct subcategories of readers: literary reviewers and Freudian critics, towards whom he adopts a provocative and ironic attitude:

21. Most people will not even mind having missed all this, well-wishers will bring their own symbols and mobiles, and portable radios, to my little party; ironists will point out the fatal fatuity of my explications in this forward and advise me to have footnotes next time. (*Bend Sinister*, 10)

There is an important difference between the nature of the intersubjective relation that Fielding and Hawthorne enact in their prefaces with their audiences, and the one enacted by Nabokov. The former adopt an ethos of modesty since they do not speak down to a subcategory of their would-be readers; they are not being ironic nor categorical in their statements and they often use hedges to voice what they think:

22. But to illustrate all this by another science, in which, perhaps, we shall see the distinction more clearly and plainly: let us examine the works of a comic history-painter, with those performances which the Italians call Caricatura, where we shall find the greatest excellence of the former to consist in the exactest copy of nature. (*Joseph Andrews*, 6)

Fielding uses inclusive engagement markers (we, let us) and the adverb *perhaps* so as not to distance himself from his would-be readers or keep them at bay. These engagement markers indicate that the interpersonal relationship he enacts with them is characterized by proximity and benevolence and they reinforce the ethos of modesty he carves out in his prefaces.

Likewise, Hawthorne adopts the ethos of modesty by appealing to his readers' indulgence:

23. He [the author] would be glad, therefore, if--especially in the quarter to which he alludes--the book may be read strictly as a romance, having a great deal more to do with the clouds overhead than with any portion of the actual soil of the County of Essex. (*The House of Seven Gables*, 4)

Nabokov, however, disseminates markers of authority in his prefaces. Most of the engagement markers he employs are stated in generic and

categorical terms; constructing, thus, the ethos of the primary knower who asserts his own interpretation of his novels and defends them against any undesirable criticism or appropriation.

## 2.2. The interactive resources

The interactive resources indicate how the writers organise their texts and signal the elements and the themes that call for more explanations or need to be made explicit. These metadiscursive elements are meant, therefore, to guide the reader to what needs to be recovered from the text either because it may be overlooked or misunderstood.

Of the five interactive resources of Hyland's classification listed in Table 1, two play a crucial role in the prefaces analysed in this paper, even if they are unevenly distributed: *code glosses* and *endophoric* markers. I will show that these two types of resources have converging exegetical ends.

### *i-Code glosses*

Writers usually use code glosses to make sure that their readers understand their intended message; they provide additional information by restating, rewording, elaborating, or explaining some of the propositional contents of the "central" text. In Fielding's, Hawthorne's and Nabokov's prefaces, code glosses are genuine exegetic commentaries rather than mere clarifications and explanations of what might otherwise be ambiguous or fuzzy. Code glosses, as we can see below, are often long stretches of sentences:

24. In the ~~d~~iction I think, burlesque itself may be sometimes admitted; of which many instances will occur in this work, as in the ~~d~~escription of the battles, and some other places not necessary to be pointed out to the classical reader; for whose entertainment those parodies or burlesque imitations are chiefly calculated. (*Joseph Andrews*, 7)
25. The point of view in which this tale comes under the romantic definition lies in the attempt to connect a by-gone time with the very present that is flitting away from us. (*The House of Seven Gables*, 6)
26. The Russian title of this novel is Zashchita Luzhina, which means "the Luzhin defense" and refers to a chess defense supposedly invented by my creature, Grandmaster Luzhin: the name rhymes with "illusion" if pronounced thickly enough to deepen the "u" into "oo". (*The Defense*, 7)
27. The oblong pool, shaped like a cell that is about to divide, reappears subthematically throughout the novel... The puddle thus kindled and rekindled in Krug's mind remains linked up with the image of his wife... (*Benjamin Sinister*, 8)

The writers gloss on what they consider their would-be readers cannot know; they tell them why the novel belongs to the burlesque genre (Fielding) or to the romance genre (Hawthorne). Not only does Nabokov inform his readers about the original title of his novel, he also alludes or overtly refers to the fundamental themes and purposes of his novels.

Code glosses in these prefaces reinforce, once again, the status of the three authors, as being the ultimate legitimate critics of their own novels. The exegetic comments conveyed by these metadiscursive elements are meant to help the readers to classify the novels they are about to read and/or identify their main topics and purposes. At the same time, they bolster the status of the novelists as the primary knowers of their own works and tend to shape the readers' horizon of expectations.

### *iii- Endophoric markers*

They typically help the reader recover the writer's intentions by referring to other parts of the text where the same meaning is available or salient. I have not found any endophoric markers in Fielding's and Hawthorne's prefaces, but there quite a few in Nabokov's. They orient the readers' attention to different chapters of the novels where the topics and themes Nabokov talks about are elaborated and to be found:

28. The retrospective theme begun in Chapter Four shades now into the image of Luzhin's father, whose own past is taken up in Chapter Five... (*The Defense*, 9)
29. Ember, for instance, in Chapter Seven, gives his friends a sample of the three lines of the Hamlet soliloquy (Act III, Scene, I) ... He follows this up with a Russian version of part of the Queen's speech in Act IV, Scene VII... and a splendid Russian rendering of the prose passage in Act III, Scene II... (*Bend Sinister*, 9)

## **3. Metadiscourse resources and reading positions**

As we have seen in the previous section, metadiscourse is fundamentally text-oriented, because it is a comment and or a reflexion on the central text, and it is participant-oriented as well, because it addresses the potential and the ideal reader and it enables the writers to position themselves towards their "central" text and to their would-be readers. The three authors draw on these two intertwined aspects of metadiscourse to appraise their novels, to interact with their potential readers, to enlist them and, most importantly, to assign them particular reading positions. Therefore, we need to investigate how the interpersonal relations enacted by the metadiscursive materials of the prefaces have a hermeneutical bearing and how they carve out and impose particular reading positions from which readers are enticed or

summoned to assess the novels introduced by the six prefaces. This hermeneutical dimension of the metadiscursive material of the prefaces can be analyzed in the light of Ricoeur's textual hermeneutics. According to Ricoeur (1998, 137-41), a text's meaning potential cannot and should not be reduced, fossilized and "encapsulated" in the author's presumed and original intentions; a text is a meaning potential that is always appropriated by each reader. Each appropriation inscribes the text in a living communication and opens up the possibility of actualizing its meaning potential differently; the world of the text combines with the world of the reader and forms something new.

The analysis of the metadiscursive elements of the six prefaces shows that each author resists, in his own way, the would-be readers' potential subjective appropriation of his work, either, by suggesting or imposing a preferred interpretation. Therefore, we can say that the metadiscursive devices of the prefaces fulfil both an interpersonal function and a hermeneutical one as well. They are not only concerned with thematic and topical managements, helping the readers navigate the meaning potential of the novels they introduce, they also instruct, explicitly and unwittingly, on how to read them.

Indeed, the three authors take a stance towards their novels and engage in interpersonal activity with their would-be readers with a particular perlocutionary aim in mind: to prepare the reception of their central texts and predispose their readers favourably. This perlocutionary goal has a hermeneutical dimension to it. If Fielding and Hawthorne seek to achieve this goal by adopting an ethos of modesty and by establishing a balanced relationship with their readers, Nabokov, however, does not seem to.

It goes without saying that the chief purpose behind these two antagonistic interpersonal relationships – Fielding and Hawthorne's on the one hand and Nabokov's on the other – is to bring to the fore the reading positions each novelist embeds and makes available to his would-be readers. Fielding and Hawthorne do not aim at overtly telling their readers how they should interpret their novels, they seem to leave them a fair amount of freedom; the two novelists suggest or orient the readers' attention to possible interpretations, without enjoining them to take what they say for granted. Fielding and Hawthorne try to narrow the gap between themselves and their readers; they do not try to coax them into a peaceful relationship, to win them over or to make them involved. Nabokov refuses explicitly to do this; he even delights in poking fun at a subcategory of his readers (literary reviewers and, most of all, psychoanalysts) while distancing himself from them by being sarcastic and often haughty.

Fielding and Hawthorne seem to be predisposed to accept potentially divergent readings of their novels; they do not assign a unique reading position which their would-be readers are summoned to occupy. They use the interpersonal potential of the metadiscursive elements in their prefaces to establish what might seem to be a genuine non-coercive communication (Sell, 2000). They allude to possible interpretations of their novels, without imposing their own; neither author openly guides or ordains the readers towards a unique way of accessing and understanding his novels. This non-coercive communication reinforces and is reinforced by the ethos of modesty they both adopt; they repeatedly make it clear that they want to forge a bond with their readers, based on solidarity and empathy. The latter are co-opted as co-constructors of the meanings of their novels; they invite them to participate in a harmonious dialogue between equals.

Nabokov, on the contrary, does impose a 'hermeneutic closure' because his ultimate aim is to shield his novels from undesired interpretations. He keeps undervaluing and warding off the responses of potential here-and-now readers and he shores up the idea that the significance of his novels is defined by, and confined to, his original intentions. Indeed, Nabokov does not allow his readers to add new interpretations to his own and to appropriate his novels without submitting to his guardianship; he seeks to prevent his novels from escaping his own horizon. In sum, he refuses to let his readers follow different paths from his own to traverse the novels, as though he feared being expropriated from the meanings of his novels.

Nabokov imposes an unbalanced relationship with his readers and he does not even pretend to be trying to hook or bring them in; he only enlists them to support the ratification of what he says and states, expecting them to acquiesce to his judgments and his readings. He carves out the solitary figure of an author whose work cannot be fully and adequately understood and appreciated by anyone else but himself. His novels cannot be affiliated to any existing tradition and no reader or critic can understand his novels as intimately and as subtly as he does. Nabokov creates a textual persona that incarnates and enacts authority and superior knowledge; readers are convoked and invoked as consenting recipients of the words of the Oracle: the author himself:

30. It is unnecessary to enlarge, in this elementary Foreword, on the more complex aspects of my chessmen and lines of play. (*The Defense*, 10)
31. Most people will not even mind having missed all this, well-wishers will bring their own symbols and mobiles, and portable radios, to my little party; ironists will point out the fatal fatuity of my explications in this forward and advise me to have footnotes next time. In the long run,



however, it is only the author's satisfaction that counts. (*Bend Sinister*, 10)

In these examples, Nabokov once again shows his potential readers that he is the only authority that can decide what is important or not, what is worthy of dealing with, and what should be known before reading the two novels in question.

Unlike Nabokov's, Fielding and Hawthorne's guidance is not overt; yet what we have in these prefaces is a difference of degree, not of nature. Indeed, even when they seem to be only making the story lines of the novels more explicit or when they specify what literary genres the novels belong to, the writers use the metadiscursive elements of the six prefaces indicate thematic prominence and give more salience to a particular interpretation each time: the author's. Therefore, the framing movement enacted by the metadiscursive elements in each preface marks and delimits an interpretative space and the author's interpretation, be it stated straightforwardly or obliquely, is presented as the legitimate and acceptable one.

In the six prefaces, the metadiscursive devices resorted to should be considered as framing moves that shape the potential readers' horizon of expectations and cue them about how to read the novels. They contribute to establish what each author deems to be the only acceptable interpretive frame, each of them in his own way. The authors hollow out a limited set of reading positions for readers to occupy; reducing, more or less, drastically the potentially unlimited series of possible reading positions. They expect their would-be readers to assume these places as well as the interpretive stances the authors adopt. Whether carried out overtly or unwittingly, the authors invite their would-be readers to comply with the assessments and evaluations proposed and imposed by themselves. Each time, these assessments and evaluations highlighted by the metadiscursive material of the novels is a passageway to the 'real' meanings of the novels.

## Conclusion

The aim behind analysing metadiscourse and stance markers in the six prefaces is to shed light on what metadiscourse material is used by three authors from different centuries in their prefaces, what goals it achieves and what differences and similarities stand out. The analysis of metadiscourse highlights the triangular relationship that exists between the author, the text and the audience; this relationship is the cornerstone of the prefaces in spite of the temporal distance that separates them. Another permanent characteristic of this triangular relationship is that it is unbalanced; indeed,

the metadiscursive material signals the author's intrusion into the ongoing text and indicates how he assigns particular reading positions to his readers. It highlights how the preface writer takes a stance towards the text introduced by the preface as much as towards the readers themselves. Finally, the analysis has also shown that even if it is not explicitly related to the propositional content of the text, metadiscourse material always plays a crucial role in negotiating meaning, controlling and sign-posting it.

I have shown that each of the three novelists activates and exploits the generic characteristics of the preface and interpersonal potentialities of metadiscourse to appraise their own novels, to construct a particular textual persona, to establish an interpersonal relation with their would-be readers so as to prepare the reception of their novels. I have shown that the metadiscourse devices employed in these prefaces play a crucial role, not so much in the organisation of the texts, as in engaging readers and steering them towards the preferable interpretations of the "central" text. With the help of the metadiscursive elements – which are text-oriented and participant oriented –, the three novelists expound what they consider as the true meanings of novels and elucidate what readers may miss or misread.

The analysis of the metadiscursive devices lays bare the discursive strategies by means of which readers are enlisted and guided to comply with what the authors say and claim. The differences between the six prefaces are not diachronic since they lie in the uses each author makes of these metadiscursive elements and of the generic characteristics of the preface. I have shown that of each of the three novelists suggests, or bluntly imposes, a canonical reading of his novels. Fielding and Hawthorne try to coax their readers and to ingratiate themselves with them by setting a seemingly balanced interpersonal relationship and by adopting an ethos of modesty. Nabokov, on the contrary, overtly seeks throughout his prefaces, to tame and intimidate his readers into a consenting reading position.

Overall, the analysis of the prefaces of my corpus, although extremely limited, has not displayed any important differences as far as the interpersonal dimension of metadiscourse and its perlocutionary goals are concerned. A larger corpus may further shed light on what is genre-specific and what idiosyncratic. However, my analysis has shown how the three novelists, each in his own way, endeavours to construct a textual person that is both himself and his double: a self-critic and a self-apologist. The metadiscourse material of the prefaces is employed to open or to contract the dialogic space of interaction with would-be readers; the authors indicate if they are willing to accept different readings and appropriations of their novels or if they pre-emptively disqualify undesired readings.

## Bibliography

- Ädel, Annelie. 2006 *Metadiscourse in L1 and L2 English*. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Crismore, Avon. 1989. *Talking with Readers: Metadiscourse as Rhetorical Act*. New York: Peter Lang.
- Fielding, Henry. 1998. *Joseph Andrews*. London, Penguin Classics.
- Fielding, Henry. *Tom Jones*. Oxford University Press, 1996, (first published in 1749).
- Genette, Gérard. *Seuils*, Paris : éd. du Seuil, 1987.
- Hawthorne, Nathaniel. 1850. *The scarlet letter*, London, Penguin Classics.
- Hawthorne, Nathaniel. 2013. *The House of the Seven Gables*. London, Penguin Classics.
- Hyland, Ken; Tse, Polly. 2004. "Metadiscourse in Academic writing: A Reappraisal", *Applied Linguistics*: 25/2; 156-177.
- Hyland, Ken. 2005 *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Hyland, Ken. 2009. "Metadiscourse: mapping interactions in academic writing", *Nordic Journal of English Studies*, Vol.9 No.2, 125-143.
- Vande Kopple, William. 1985. "Some Exploratory Discourse on Metadiscourse College" *Composition and Communication*, Vol. 36, No. 1, 82-93.
- Makkonen-Craig, Henna. 2011. "Connecting with the reader: participant-oriented metadiscourse in newspapers", *Text and Talk*, 31 - 6, 683-704.
- Nabokov, Vladimir. *Bend Sinister*. 1974. Hammondsport: Penguin Books.
- Nabokov, Vladimir. 1970. *The Defense: a novel*. New York: Capricorn Books.
- Ricoeur, Paul. 1998. « Qu'est-ce qu'un texte ? », *Du texte à l'action. Essai d'herméneutique II*. Paris, Seuil, Points essais, 137-141.
- Saki, Mohamed. 2013. "Intersubjective positioning and community-making in E.E. Cummings' preface to his *Collected Poems (1923-1958)*" in *The Ethics of Literary Communication*, edited by Roger D. Sell et al, Amsterdam, John Benjamins, 47-60.
- Sell, Roger. 2000. *Literature as communication: The Foundations of Mediating Criticism*, Amsterdam/Philadelphia: John Benjamins.



## CHAPTER SEVEN

### LES DEFINITIONS DU MOT « CANCER » DANS LES DICTIONNAIRES ET LES TRAITES MEDICAUX : APPROCHE DIACHRONIQUE

### DEFINITIONS OF THE WORD “CANCER” IN DICTIONARIES AND MEDICAL WORKS: A DIACHRONIC APPROACH

GHISLAINE ROLLAND-LOZACHMEUR

#### **Abstract**

Cancer is a serious illness that has caused fear since ancient times. Authors use various metaphors to define it, and dictionaries and medical works on cancer help transmit the fear generated by the word. The objective of this article is to analyse the definitions of the word “cancer” in language dictionaries and medical works, from the 18<sup>th</sup> century up to today’s health advice and information websites. The study will take into account the distribution of the word in relation to its syntactic functions. The analysis of the different definition practices reveals a tendency to adopt a convenient and condensed defining form as a writing process, with constant repetition of the same descriptions and syntagmatic combinations in monographs devoted to cancer, and shows how the disease is associated with a lexicon of combat.

## Introduction

Les études de discours en diachronie occupent une place importante dans les travaux menés en linguistique. Comme ceux qui les pratiquent, les langues ont une histoire, indispensable à la compréhension de la vie sociale et de la culture ; la linguistique historique s'inscrit dans cette recherche. Progressivement, elle a intégré des considérations sur l'évolution des sens et le contenu des unités du lexique. Travailler en diachronie permet d'étudier les faits linguistiques selon l'axe des successivités temporelles et d'expliquer les changements survenus dans la langue.

Cette interpénétration des domaines « histoire/linguistique », mise en évidence par Robin (1973), n'est pas sans présenter des difficultés, notamment en lexicographie (Branca 1988, 47 ; Guilhaumou 1989, 12).

Une étude centrée sur les mots doit tenir compte de cette perspective. Pour Eluerd (2000, 63), « [l]es mots ne sont pas plus de simples reflets des faits sociaux et historiques qu'ils ne sont des reflets du monde » ; les mots sont des témoins. Il rappelle « l'importance du ' flux de vie ' (ou forme de vie) selon Wittgenstein (1954) qui donne leur signification aux énoncés linguistiques ».

L'étude du mot « cancer » est ici envisagée dans la continuité de nos travaux sur l'analyse du discours en cancérologie (Rolland-Lozachmeur 2015). En effet le mot cancer, tout comme la maladie qu'il représente, peut se suivre en diachronie. Il n'est jamais sorti de la langue comme d'autres mots appartenant au lexique héréditaire (Perret 1998, 103–115). Si l'on parcourt ainsi les dictionnaires de langue et les dictionnaires médicaux anciens, il ressort des descriptions monographiques qui lui sont consacrées que le cancer est décrit comme une maladie grave qui fait peur, qui terrifie. On a d'ailleurs recours à la métaphore pour la dénommer, comme si l'image utilisée permettait d'en éloigner la crainte (Oliveira 2009). Cette maladie est connue depuis la haute Antiquité (1600 av. J.-C.) et déjà le sein était considéré comme une cible privilégiée. Rouéssé (2011, 16) explique :

Depuis qu'il est connu, il est regardé comme une maladie aussi affreuse que bien souvent fatale malgré les traitements : c'est une prolifération anarchique, tumorale ou ulcérate s'étendant inexorablement. « *Le cancer se jette plus volontiers aux mamelles à cause de leur faiblesse, qu'elles sont vuides (sic) et fort délicates.* »

De fait, le cancer est un symbole de mort ; le grand public le redoute. Pour mieux comprendre l'impact du mot sur les locuteurs, il s'agit d'observer comment ses définitions évoluent dans les dictionnaires de langue et de science, dans les traités de médecine, puis dans le discours



contemporain sur les sites Internet. Suivre le mot en diachronie dans les pratiques définitoires depuis le XVII<sup>e</sup> siècle et enregistrer les sens par une analyse distributionnelle permettent de saisir l'évolution des représentations sociales de la maladie.

Le présent article pose, dans une première partie, les caractères du genre définitoire et ses marques linguistiques. Puis l'évolution diachronique des définitions du mot « cancer » est exposée dans la deuxième partie. Enfin, la troisième partie isole les caractères morphosyntaxiques et lexicaux constants dans le corpus étudié.

## 1. La pratique définitoire

### 1.1 La définition

La définition du dictionnaire répond à une démarche langagière et proprement philologique, respectueuse des variations que l'usage social produit pour une forme linguistique donnée, généralement lexicale. Cette démarche aboutit à un article de dictionnaire de langue, qui devient un ensemble de définitions et de gloses appuyé sur des exemples (Rey 1990, 13) :

Définir un mot ou un signe stable dans un langage langue naturelle ou non , c'est en effet, soit mettre en rapport ce signe avec d'autres, au même niveau sémiotique (et on parle par exemple de synonymie périphrastique, ce qui convient à la définition du dictionnaire), soit avec les signes d'un métalangage construit, ce qui peut être le cas en sciences, parfois en philosophie ou dans n'importe quelle théorie soucieuse de sa terminologie.

Pour le chercheur, l'activité définitoire présente une forme brève, concise qui est utile et significative à plus d'un titre : elle rassemble l'information et renseigne sur les intentions de l'énonciateur, sur les débats d'une époque donnée.

Elle est aussi une forme textuelle qui fait partie des stratégies d'écriture, se prêtant au jeu de l'argumentaire et prenant au fil du temps historique des aspects variés : anecdote, parodie de dialogues, argumentations pseudo-scientifiques. On peut ainsi présenter des informations sous la forme définitoire. D'ailleurs, pour Quemada (1968, 418), lexicographe, on a bien une réalité correspondant à cette définition :

[F]ait sociologique autant que linguistique ou scientifique, le dictionnaire est avant tout un catalogue de vocables destinés à donner accès, de manière

commode, à des informations dont le nombre et la nature sont laissés à l'appréciation de l'auteur.

Au plan lexicographique, le dictionnaire de langue propose une stricte description du lexique avec le mode d'emploi de chacun des mots. Il rassemble des données concernant la forme, le contenu sémantique des mots, leurs modalités d'utilisation, la situation des mots à l'intérieur du système linguistique décrit, c'est-à-dire l'inclusion des mots dans les séries lexicales (morpho grammaticales, synonymiques, antonymiques, analogiques) et dans les champs thématiques. Le dictionnaire encyclopédique (Furetière, Trévoux, Encyclopédie de Diderot, Larousse) donne des renseignements sur la chose désignée par le mot : son utilisation, son origine et sa place dans la culture de la communauté. Les dictionnaires étant plus ou moins hétérogènes, il y a ainsi continuité du dictionnaire de langue à l'encyclopédie.

## 1.2 Les marques linguistiques de la définition

L'auteur d'un dictionnaire recourt à un métalangage dont Quemada (1968, 417-418) a dégagé des traits distinctifs :

- le définisseur générique : la métalangue lexicographique possède son lexique ; des mots comme « espèce », « sorte », « manière » « qualité », des symboles abrégatifs, autant d'opérations syntaxiques, de caractères typographiques propres, qui expriment l'essence et la nature de la chose désignée par le moyen d'une indication classificatoire, le genre.
- les marques spécifiques : une ou plusieurs notations caractéristiques complètent le définisseur générique variant par leur nombre, leur nature et leurs emplois qui sont d'ordre descriptif, causal, génétique, fonctionnel, modal, explicatif ou à visée finale. Elles oscillent entre rattachement notionnel (parenté, appartenance), rapports étymologiques et rapports analogiques.
- les termes de liaison assurent la cohérence syntaxique de la définition, outre les mots outils, par le moyen de groupes nominaux, de relatives.

Evidemment, l'énoncé peut ne pas prendre la forme attendue par ce schéma formel. Dès Furetière, le dictionnaire renseigne sur le monde, les choses. Cette tendance s'accroît avec la série des Trévoux (1704-1771) et avec l'*Encyclopédie* qui classe des connaissances. Le procédé est commode car il évite de se poser des problèmes de plan et offre un support visuel par ordre alphabétique qui permet de trouver les matières plus rapidement. On

distingue ainsi plusieurs types de définitions parmi lesquelles l'encyclopédique centrée sur la chose désignée par le mot (nomenclature nominale), descriptive avec un véritable développement (Quemada 1968, 411), et la grammaticale inspirée des philosophes de Port-Royal, pratiquée dans les dictionnaires de langue, normatifs ou d'usage, décrivant le matériel et le système de la langue, les particularités linguistiques du signe. La définition ne dit pas tant l'essence de la chose que la représentation que s'en font les hommes : « Toute définition du mot est avant tout « la déclaration de l'usage et des idées que les hommes y ont appliquées » (Chaurand 1990). Cela dit, la définition est aussi, depuis Aristote, considérée comme un moyen d'argumenter (Perelman & Olbrechts-Tyteca 1988, 288)

### 1.3 Rôle de la définition au-delà du métalangage

Un autre enjeu de la définition est d'établir une relation étroite entre sémantique et syntaxe. Pour analyser les propriétés externes du mot, ses distributions, c'est-à-dire son comportement dans le texte, on observe son fonctionnement syntaxique : le verbe, dont le définisseur « être », au rôle argumentatif non négligeable, est au premier plan, les prépositions, les compléments et leur nature. L'analyse distributionnelle associée à la syntaxe joue un rôle primordial en tant que révélateur du sens : on cherche les traits de sélection (contraintes lexicales) et la relation isotopique introduite entre les mots de différentes classes grammaticales. L'activité définitoire sollicite également la méthode des champs actanciels, pour repérer dans la structure phrastique les actants du verbe, leurs dénominations, leurs qualifications et les dérivés sémantiques du verbe. Le mot étant un domaine d'investigation efficace, il faut tenir compte, notamment, de sa disposition en chaîne dans la phrase, de son contexte immédiat, son sens résultant précisément des combinaisons auxquelles il donne lieu (Benveniste 1966, 227) :

Le sens d'un mot consistera dans sa capacité d'être l'intégrant d'un syntagme particulier et de remplir une fonction propositionnelle. Ce qu'on appelle la polysémie n'est que la somme institutionnalisée, si l'on peut dire, de ces valeurs contextuelles, toujours instantanées, aptes continuellement à s'enrichir, à disparaître, bref, sans permanence, sans valeur constante.

Cette « polysémie » peut se réaliser au sein de collocations qu'une lexie forme avec d'autres lexies de la langue sur l'axe syntagmatique. Sinclair et al (2005, 10) précisent : « collocation is the co-occurrence of two items in a text within a specified environment ». La construction du sens obéit au principe de compositionnalité sémantique selon lequel « le sens d'un énoncé

est la résultante de la composition du sens des éléments qui le constituent » (Polguère 2003, 134). Le sens est contrôlé par la base de la collocation qui sélectionne un collocatif. Ce dernier peut être un modificateur de la base (adjectif épithète pour le nom ou adverbe pour le verbe ou l'adjectif) ou un verbe support (collocatif verbal qui verbalise une base nominale comme si elle était un verbe).

Ainsi dans l'environnement du mot cancer quels verbes, quels groupes nominaux ou adjectivaux trouve-t-on et quels sont, notamment, les champs actanciels de ces verbes ?

## 2. Les définitions de « cancer » en sémantique diachronique

Dans cette partie, nous présentons les caractères des définitions du mot « cancer », dans un découpage qui nous a permis de distinguer quatre étapes dans leur élaboration. Pour cela, nous avons systématiquement relevé les monographies du cancer présentes dans cinq dictionnaires de l'époque classique (Richelet 1680, *Dictionnaire de l'Académie française* première édition de 1694 et quatrième édition de 1762 ; Furetière 1690 ; *Encyclopédie* de Diderot 1751), un traité sur le cancer de Dupré de Lisle (1774), un dictionnaire de médecine D'Adelon (1834), des dictionnaires du XX<sup>e</sup> siècle (le Robert historique 1992, le *Dictionnaire de l'Académie française* 8<sup>e</sup> édition 1932–1935, et 9<sup>e</sup> édition (version en cours), *Larousse médical*, *Trésor de la Langue Française* informatisé), deux traités médicaux (Tripodi 2009 ; Paillaud & Clément-Hryniewicz 2008), les sites de santé Internet (AlloDocteur.fr.France5, Docteurcliv, Vulgarismedical, doctissimo, passeportsante).

### 2.1. Le mot « cancer » dans les dictionnaires des XVII<sup>e</sup> – XVIII<sup>e</sup> : de la description à la subjectivité

Au début du XVII<sup>e</sup>, le cancer terrifie même si les données du XVIII<sup>e</sup> siècle indiquant le taux de décès par mort du cancer paraissent faibles en pourcentage (8 %). Le public, médical ou non, a alors l'impression que le cancer devient de plus en plus fréquent (Rouëssé 2011). Aussi en répertoriant les définitions du mot cancer dans les principaux dictionnaires de cette période, dont nous donnons quelques exemples en orthographe non rectifiée, peut-on évaluer le poids de cette maladie dans les mentalités :

- (1) -Richelet (1680)<sup>1</sup>: « Cancer, s.m. Tumeur impure, maligne, ronde & inégale, qui est au commencement sans douleur & qui est engendrée d'une humeur atrabilaire.  
Cancer, s.m. Un des douze signes célestes, qu'on appelle aussi Ecrevisse. »
- (2) -Dictionnaire de l'Académie (1694, 1<sup>re</sup> édition)<sup>2</sup>: Cancer. S. m. Espece de tumeur maligne qui dégenere en ulcere presque incurable. Il luy est venu un cancer du sein. »
- (3) -Dictionnaire universel de Furetière (1690)<sup>3</sup>:« Cancer.S.m.Terme de médecine. C'est une maladie qui vient dans les chairs, & qui leqs<sup>4</sup> mange petit à petit comme une espece de cangreine. C'est une tumeur dure, inégale, raboteuse, ronde & immobile, de couleur cendrée, livide ou plombine, environnée de plusieurs veines apparentes & tortués, pleines d'un sang melancolique & limoneux, qui ressemblent au poisson appelée *cancer* ou écrevisse. Elle commence sans douleur, & paroist d'abord comme un pois chiche ou une petite noisette ; mais elle croist assez vite, & devient fort douloureuse. Les *cancers* viennent aux parties glanduleuses & lâches, comme aux mamelles & aux emonctoires. En grec *karkinos*, qui signifie aussi *écrevisse*. Ce mal a grand rapport avec cette sorte de poisson, en ce que quand une fois il a pris pied dans un corps, il est presque impossible de l'en chasser, de même qu'il est difficile d'arracher des pinces de l'écrevisse ce qu'elle a une fois attrapé»
- (4) -Dictionnaire de l'Académie 4<sup>e</sup> édition 1762<sup>5</sup> : « Espèce de tumeur maligne qui dégénère en ulcère, principalement au sein. *Une femme qui a un cancer au sein. Arracher un cancer. Extirper un cancer. Elle a le sein tout rongé d'un cancer.* En Chirurgie, *Cancer de Galien* est une espèce de bandage. »
- (5) -*Encyclopédie de Diderot* <sup>6</sup>: « Cancer : terme de chirurgie. est une tumeur dure, inégale, raboteufe, & de couleur cendrée ou livide, environnée tout-à-tour de plusieurs veines diffendues & gonflées d'un sang noir & limoneux - située à quelque partie glanduleufe ainfi appelée à ce que quelques-uns prétendent, parce qu'elle est à-peu-près de la figure d'une écrevisse ou, à ce que disent d'autres, parce que semblable à l'écrevisse elle ne quitte pas prise quand une fois elle s'est jettée sur une partie. »

● On distingue deux pratiques différentes : le dictionnaire de langue vise les sémantismes lexicaux et les significations dans l'usage alors que le

<sup>1</sup> <http://gallica.bnf.fr/ark:/12148/bpt6k509323>

<sup>2</sup> <http://artfl.atilf.fr/dictionnaires/ACADEMIE/PREMIERE/premiere.fr.html>

<sup>3</sup> <http://gallica.bnf.fr/ark:/12148/bpt6k50614b>

<sup>4</sup> Orthographe non rectifiée du XVII<sup>e</sup> siècle : qui mange « les chairs »

<sup>5</sup> <http://artfl.atilf.fr/dictionnaires/ACADEMIE/QUATRIEME/quatrieme.fr.html>

<sup>6</sup> <http://gallica.bnf.fr/ark:/12148/bpt6k505351>



dictionnaire « raisonné », concept de la logique et de la philosophie, tel l'*Encyclopédie* vise les notions et leur incarnation par des termes, ces derniers étant confondus avec les mots et les syntagmes qui les représentent (Rey 1990, 17–18).

Dans les définitions du XVII<sup>e</sup> et du XVIII<sup>e</sup> du mot cancer, les marques spécifiques sont en relation avec la théorie des humeurs, ce que rappellent les termes « humeur » et « sang mélancolique ». Furetière entérine ainsi dans sa définition des connaissances sur le cancer déjà répertoriées depuis l'antiquité, notamment la proximité descriptive avec l'écrevisse (ailleurs le crabe), le caractère intrusif et l'évolution inexorable de la maladie et le témoignage des interrogations médicales de son époque. Ce sont des dictionnaires de « choses » avec des définitions très descriptives. Chez Furetière, comme d'une façon générale, les marqueurs génériques sont métalinguistiques ou hyperonymiques : « terme de », « espèce » ou « maladie ». Les liaisons syntaxiques sont des présentatifs, des déterminants démonstratifs (« c'est », « ce » [« ce mal »]), des expansions du nom tête (« espèce de tumeur »), des adjectifs (« maligne »), des compléments de détermination du nom et de l'adjectif, des relatives (« qui dégénère »). En outre, à côté d'« écrevisse », on relève des métaphores peu fréquentes comme « pois chiche » ou « petite noisette ».

Avec Dupré de Lisle (1774) au XVIII<sup>e</sup> siècle, on assiste à une rupture dans la définition rencontrée jusqu'alors. C'est avant tout un médecin qui expose les recherches théoriques et expérimentales sur le cancer, les causes et les nouveautés thérapeutiques et qui cherche à préciser la définition. Il qualifie la maladie, dès le début de l'ouvrage, de « putréfaction cancéreuse », de « maladie des plus redoutables qui puisse le plus intéresser l'humanité » (*ibidem* xiv). Il la présente comme un fléau, une maladie qui fait peur et qui met en difficulté la médecine :

Galien est le premier qui donna le nom de Cancer à cette tumeur que nous traitons : il voulut l'appeler ainsi, par rapport à une espèce de ressemblance qu'elle a avec l'écrevisse. Comme cet animal présente plusieurs pattes de part & d'autre, de même on voit partir de la tumeur chancreuse une infinité de ramifications pleines d'un sang noir. (de Lisle 1774 : 4)

Du fait de la comparaison avec l'écrevisse, la définition est très subjective ; elle s'appuie sur les adjectifs « redoutable » et « hideuse », comme dans les exemples suivants :

Le Cancer n'est autre chose qu'une tumeur très-irrégulière qui résulte de la conversion des parties squarreuses en une substance hideuse, très susceptible d'un singulier accroissement ainsi que d'ulcération. (*ibidem* 13)



La tumeur qu'on appelle taupe, & qui se forme à la tête, n'est qu'un follicule membraneux, rempli d'une espece d'humeur épaisse. (*ibidem* : 38)

Dans cette perspective, l'objectif semble être celui de faire le point sur l'état des connaissances sur le cancer et de donner un état descriptif de la situation : « Nous allons nous occuper du vice cancéreux. L'on sait qu'il est un des plus redoutables fléaux de l'humanité, & l'écueil de l'art » (*ibidem* ix). L'intérêt de ce traité est de dresser une classification des types de cancers en fonction de leur évolution, classification absente des dictionnaires :

Nous avons assigné trois caracteres très-distincts les uns des autres dans les tumeurs dégénéralbles en cancer. Le premier, par rapport à sa dureté, a été connu sous le nom de Squirre faux ou légitime. Le second prend le nom de Cancer occulte, commençant ou invétéré. Le troisième enfin a été appellé Cancer ouvert ou ulcéré, & par les Anciens Carcinome. *ibidem*, p.119).

Les énoncés définitoires permettent ainsi de délimiter le territoire du connu, les domaines de connaissances partagées, même si l'auteur suggère que ces connaissances sont soumises à l'incertitude.

La description du cancer, à l'époque classique, se fait donc avec des qualificatifs qui supposent une évaluation en termes axiologiques bons/mauvais. Parfois, les analyses des causes sont très proches des analyses contemporaines : rôle du stress ou de l'abus de l'alcool dans la maladie. Par exemple, Dupré de Lisle explique le cancer avec des termes techniques mais en même temps, ce qui marque une évolution, y mêle des considérations morales sur l'évaluation des effets ressentis, les comportements contraires à la norme : « abus » ou « trop grand usage du vin ».

## 2.2. Les dictionnaires de médecine au XIX<sup>e</sup> : importance des observations techniques

Le dictionnaire d'Adelon (Adelon et Béchara 1822, Tome 4<sup>e</sup>) utilise le mot squirre (ou squirre), « Forme de cancer de consistance dure du fait de la prédominance d'une sclérose avec rétraction des tissus » Les modalités très fortes et les termes plus techniques caractérisent ses définitions :

CANCER, s. m., cancer. Il est extrêmement difficile de donner une définition du cancer qui puisse faire connaître la nature intime ou l'essence de cette maladie. Néanmoins dans l'état actuel de la science, et d'après de nombreux faits observés sous les rapports cliniques et anatomiques, nous croyons pouvoir considérer le squirre comme une induration produite par l'exhalation et le séjour d'une matière concrescible dans les alvéoles de nos

*tissus, par suite d'une irritation; les ulcères carcinomateux la matière cérébriforme, le fongus hématoïde, le sarcome médullaire, etc., comme la dégénérescence produite par une inflammation secondaire qui détermine la fonte des tissus qui existaient déjà dans une condition morbide... (ibidem)*

Dans l'entrée « cancer » de l'édition de 1834, Adelon approfondit l'origine et l'utilisation du terme :

Le mot latin cancer signifie crabe, cancre et correspond au mot karkinos des Grecs, qui sert à désigner le même animal. Il est intéressant de rechercher si ce mot a été introduit dans la science par suite d'une prétendue ressemblance entre les veines dilatées qui s'écartent en rayonnant d'une tumeur cancéreuse, et les pattes d'un crabe, ou parce qu'on a cru effectivement qu'un animal dévorait les parties malades (Littre in Adelon (1834), 306)

On observe un détachement, un retrait dans la façon d'aborder la maladie : « il est intéressant », « ce qu'il importerait d'établir », « aussi est-il devenu difficile avec les structures impersonnelles et modalisantes (pouvoir + inf, le mode conditionnel) et des périphrases actanciennes passives. Si l'on s'interroge sur la motivation du mot « crabe », Adelon émet ici plusieurs hypothèses dans son ouvrage.

### 2.3. Les dictionnaires, ouvrages spécialisés du XX<sup>e</sup> siècle : le renforcement de la terminologie scientifique

(1) Robert Dictionnaire historique de la langue française (Rey 1992 (2004), 603) fournit la définition suivante :

Cancer : n.masculin vient (1372) du latin cancer, cancri (cancre, chancre, crabe) qui traduit en reprenant les sens, le grec karkinos « crabe », « chancre », également « pincés », « paires de compas ». Les deux mots appartiennent, avec le grec karkaros « dur » et le sanskrit karkata « crabe », à une même racine indo-européenne. Le mot a été introduit comme nom d'un signe du zodiaque. Le sens médical de « tumeur maligne » (1476) a donné un emploi figuré (av. 1755) et s'est spécialisé (fin XIX<sup>e</sup> s.) pour désigner le néoplasme (prolifération pathologique de cellules). Puis l'idée de « tumeur » n'étant plus essentielle, le mot a désigné tout état pathologique caractérisé par des lésions résultant d'une prolifération cellulaire non contrôlée par l'organisme (cancer des os, cancer du sang). Le cancer est à ce point considéré comme le fléau, la maladie du XX<sup>e</sup> s. par excellence (comme l'était la peste au moyen âge) que des journalistes ont récemment défini le sida comme le cancer du XXI<sup>e</sup> siècle.

(2) Dictionnaire de l'Académie 8<sup>e</sup> édition<sup>7</sup> (1932–1935) propose une définition minimale, mais reprenant les sèmes principaux comme présence d'une « tumeur », et maladie qui « dégénère » ou qui ronge :

Cancer (nm) Terme générique de Médecine par lequel on désigne toute tumeur qui dégénère en ulcère et qui ronge les tissus. Cette femme a un cancer au sein. Cancer de l'estomac, de la langue, de la vessie, etc.

(3) Dictionnaire de l'Académie 9<sup>e</sup> édition<sup>8</sup> (en cours) est plus précis :

N m. Emprunté du latin *cancer, cancri*, traduction du grec *karkinos*, « écrevisse », puis terme d'astronomie et de médecine. Pathol. Terme générique désignant une tumeur maligne constituée de cellules proliférant anormalement, qui peuvent envahir les tissus voisins et donner lieu, à distance, à des métastases. *Cancer du sein, de l'estomac, de la langue, de la vessie. Initialement réservé aux épithéliomas, le terme de cancer est couramment appliqué à toutes les variétés de tumeurs malignes.* Fig. Mal insidieux susceptible de se généraliser. *Le cancer du chômage dans les sociétés industrielles.*

● On remarque que le sens figuré s'étend. Est-ce le signe que le mot est entré dans l'usage ? Le verbe « envahir » avec la notion de « métastases » supplante « ronger ».

(4) Larousse médical<sup>9</sup>, dresse également un inventaire des types de cancer :

Cancer : Maladie qui a pour mécanisme une prolifération cellulaire anarchique, incontrôlée et incessante.

(5) - TLFi<sup>10</sup> Trésor de la langue française informatisé, permet de saisir le mot en discours littéraire et politique :

Cancer : Tumeur maligne due à une multiplication anarchique des cellules d'un tissu organique. « Elle a dix-huit ans ; il y en a cinq qu'elle est tourmentée par un horrible cancer qui lui ronge la tête. » J. De Maistre, *Les Soirées de Saint-Petersbourg*, t. 1, 1821, p. 251. *Par. Méaphore. ou au figuré.* [En parlant d'une situation, des systèmes de valeurs, de l'usage d'un pouvoir, d'une influence devenus excessifs, etc.] Détérioration lente. « Cependant le cancer de l'esclavage gagnait de proche en proche » (Michelet,

<sup>7</sup> <http://atilf.atilf.fr/academie.htm>

<sup>8</sup> <http://atilf.atilf.fr/academie9.htm>

<sup>9</sup> <http://www.larousse.fr/archives/medical/>

<sup>10</sup> <http://atilf.atilf.fr/tlf.htm>

*Introd. à l'Histoire universelle*, 1831, p. 417). « Souvent le cancer du plaisir ronge les lieux et les êtres auxquels il s'attache ... » (Mauriac, *Journal I*, 1934, p. 10), « La littérature moderne, en beaucoup de cas, est un cancer des mots ». Sartre, *Situations II*, 1948, p. 304

En dehors de la lexicographie, la thèse de médecine (Tripodi 2009) offre un exemple des traits strictement scientifiques appliqués au mot cancer sur le terrain médical :

(6) Un cancer est caractérisé par une prolifération anarchique de cellules, provoquant des tumeurs dans différents organes. Il peut s'étendre localement en envahissant les tissus voisins, et/ou se disséminer par voie lymphatique et sanguine, ce qui crée des localisations secondaires. Le développement des cancers peut être lié à des facteurs de risque environnementaux, professionnels ou héréditaires. [...] Une tumeur maligne présente plusieurs anomalies phénotypiques telles qu'une croissance excessive, un potentiel d'envahissement local et une capacité à métastaser. Ces caractéristiques apparaissent progressivement au sein de la cellule, étape par étape, associant des mutations qui réalisent l'initiation cellulaire et des mécanismes épigénétiques tels que la promotion qui favorise l'expansion clonale des cellules initiées.

En revanche, l'ouvrage spécialisé en psycho-oncologie de Paillaud & Clément-Hrynivich (2008) met l'accent sur les conséquences de la maladie dans la vie des personnes atteintes :

(7) Les mots ressemblent aux rayons X. Ils peuvent transpercer n'importe quoi. Une phrase qui, dans son contexte, est à saisir dans sa double valence car, s'il est des mots qui peuvent nous être agréables, des mots qui flattent notre ego, des mots dont nous aimons nous souvenir, il en est d'autres qui peuvent très rapidement devenir l'objet de tous nos maux et qui ne manquent pas de raviver certaines blessures, de susciter notre peur. Il est ainsi du mot cancer qui, dans son sens d'origine, désigne une tumeur maligne évoquant le plus souvent le pire, le négatif et peu le positif même si les mentalités semblent changer à ce sujet. » [...] Nombreux sont en effet les écrits, les témoignages de patients, de proches pour qui le mot « cancer » est synonyme de mort, de souffrance, de traitements longs, pénibles...<sup>11</sup>

Le mot cancer est bien un mot qui effraie. Les auteurs insistent sur cet aspect :

---

<sup>11</sup> *Ibidem*

Le mot « cancer » choque. Il dérange. Il est cet intrus, étrangement inquiétant, invisible, sournois, impitoyable [...], ce mal qui vient à s'immiscer subrepticement dans la vie de certains, là où ils ne l'attendent pas ou l'attendent trop...[...] Il est ce mot, enfin, qui provoque une fracture ouverte entre un avant et un après pour les patients qui expriment leur sentiment de vivre depuis avec une épée de Damoclès au-dessus de leur tête et s'indignent de l'attitude de la société à leur égard.<sup>12</sup>

Les analyses vont dans le sens de Zorn (1977) qui, dans un registre autobiographique, raconte comment le cancer a bouleversé sa vie. Son expérience angoissante est l'objet essentiel de la mise en récit. Les mots pèsent violemment dans son écriture :

Mais le résultat de cette histoire familiale, ce qui, en tant que produit, est sorti de cette famille qui est la mienne, cette épave humaine à laquelle je dis « Je », là-dessus il me faut revenir, encore et toujours, car la connaissance de mon état de destruction me perfore sans arrêt comme une mitrailleuse. (Zorn 1977, 232)

Dans les textes étudiés ici, les adjectifs, les synonymes sont également subjectifs ou violents pour exprimer le choc ressenti.

## 2.4 Les sites Internet : la recherche de l'information exhaustive

Les sites étudiés sont les suivants : AlloDocteur.fr.France5, Docteur Clic, Vulgaris Medical, Doctissimo et Passeport Santé. Le cancer y est défini et expliqué dans des énoncés à la fois descriptifs, informatifs sur l'identification et les caractères de la maladie, et argumentatifs et didactiques quant aux causes et à la prévention.

### (1) AlloDocteur.fr France 5<sup>13</sup>

Le cancer est une tumeur maligne caractérisée par la prolifération anarchique de cellules qui se fait d'abord localement, puis au niveau des tissus avoisinants et enfin à distance (on parle de métastase). Il peut toucher tous les organes, mais les cancers les plus fréquents sont : le cancer du sein, le cancer de la prostate, le cancer du poumon, le cancer colorectal. L'origine est multifactorielle, associant des facteurs génétiques, hormonaux, environnementaux, professionnels et comportementaux (comme l'alcool, l'alimentation, le tabac et même la sédentarité pour certains cancers). Si la fréquence du cancer est en augmentation, la mortalité diminue grâce aux progrès des traitements et du dépistage précoce.

---

<sup>12</sup> *Ibidem*

<sup>13</sup> <http://www.allodocteurs.fr/maladies/cancer/>

(2) Docteur Clic Prenez votre santé en main<sup>14</sup>

« Cette rubrique vous permettra de faire le point sur le cancer. Vous pourrez ainsi mieux comprendre ce qui se passe quand on a un cancer, mais aussi comprendre un peu mieux la peur que chacun d'entre nous peut éprouver face à cette maladie, ainsi que les espoirs que les traitements actuels peuvent apporter. Le cancer est une maladie multifactorielle, c'est-à-dire dont la survenue dépend de nombreux éléments qui se surajoutent les uns aux autres. »

(3) Vulgaris Médical<sup>15</sup>

Latin : cancer, crabe.

*Le cancer est le terme désignant l'ensemble des tumeurs malignes se développant rapidement, et ayant tendance à se généraliser (métastases).*

On appelle cancérisation la transformation des cellules saines composant un tissu, en cellules néoplasiques (cancéreuses) du même type.

Tous les tissus de l'organisme sont susceptibles de subir une cancérisation ; aucun organe n'y échappe.

(4) Doctissimo<sup>16</sup>

Il existe des centaines de cancers. Si cette maladie est parmi les plus redoutées, connaît-on pour autant ses mécanismes ? [...]

Une cellule à la folie contagieuse

Si les instructions relatives à sa reproduction et à sa mort sont altérées, la cellule peut commencer à se diviser de manière incontrôlée... et devenir immortelle. [...]

La tumeur peut contrôler son ampleur et ne pas migrer vers d'autres organes, dans ce cas, elle sera dite non cancéreuse ou bénigne. Si elle envahit les organes alentour elle est dite maligne ou cancéreuse.

La cellule cancéreuse se distingue ainsi par ces deux capacités : une multiplication incontrôlée et anarchique associée à la possibilité de coloniser des tissus ou organes dont elle n'est pas originellement issue.

(5) Passeport Santé<sup>17</sup>

Le cancer est une maladie redoutée, souvent perçue comme « la pire des maladies ». Elle est la première cause de décès avant l'âge de 65 ans, au Canada et en France. De nos jours, de plus en plus de gens reçoivent un diagnostic de cancer, mais heureusement, bon nombre en guérissent. Il existe plus d'une centaine de variétés de cancer, ou tumeur maligne, qui peuvent se loger dans différents tissus et organes. Chez les gens atteints de

<sup>14</sup> <http://www.docteurclic.com/maladie/peur-du-cancer.aspx>.

<sup>15</sup> <http://www.vulgaris-medical.com/>

<sup>16</sup> <http://www.doctissimo.fr/>

<sup>17</sup> [http://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer\\_vue\\_ensemble\\_pm](http://www.passeportsante.net/fr/Maux/Problemes/Fiche.aspx?doc=cancer_vue_ensemble_pm)



cancer, certaines cellules se multiplient de façon exagérée et non contrôlée. Les gènes de ces cellules dérégulées ont subi des altérations, ou mutations. Parfois, les cellules cancéreuses envahissent les tissus environnants, ou se détachent de la tumeur d'origine et migrent vers d'autres régions du corps. Ce sont les « métastases ».

La syntaxe de ces textes se caractérise par des phrases simples et des subordonnées conjonctives, relatives et circonstancielles (comparaison, cause, hypothèse). Des interrogatives directes orientent la réflexion : « Si cette maladie est parmi les plus redoutées, connaît-on pour autant ses mécanismes ? » (citation 4, ci-dessus).

Le lexique est technique ; il est question de « métastases » (citations 1,5), de « cellules néoplasiques » (citation 3), de « dérèglement anarchique » (citation 1). On trouve des verbes qui suggèrent la progression du mal comme « migrer » (citation 4), « envahir » (citations 4,5). Dans les définitions et l'exposé des caractères, des sèmes nouveaux apparaissent avec, notamment, les termes « multifactorielle » (citation 2), « types » (citation 3), « mutation des cellules » (citation 5), « mécanismes cellulaires » (citation 4), « prolifération anarchique de cellules » (citation 1), « multiplication incontrôlée et anarchique » (citation 4). Cependant, certains auteurs peuvent aussi mentionner des expressions simples telle que « boule dans le sein ». Il semble bien que cette évolution sémantique corresponde aux attentes de la société contemporaine soucieuse de recevoir une information exhaustive et scientifique. Cette tendance rejoint l'importante fréquentation observée des sites Internet de santé et des forums sur le cancer (Rolland-Lozachmeur 2012a).

Les mots de la quantification et de l'approximation, associés à « cancer » sont plus fréquents qu'au cours des périodes précédentes : « de plus en plus reçoivent un diagnostic de cancer », « bon nombre en guérissent ». Si l'on s'intéresse à l'évolution morphologique du mot « cancer », les dérivés créés procèdent du sens médical de « cancer ». Par exemple, *cancéreux* (atteint par une tumeur maligne) a donné *anticancéreux*. *Cancérisation*, nom féminin antérieur à *cancériser* ne se répand qu'après 1920, avec *cancérisé*, puis *cancérisant* et la forme pronominale *se cancériser*. On ajoute les composés en *cancéri-*, *cancéro-* à partir de 1920 : *cancérologue*, spécialiste du cancer, puis *cancérologie*, *cancérigène* et *cancérose*. On observe ainsi des dérivés bien implantés dans l'usage comme *cancérologie* ou *carcinologie*, (étude scientifique des cancers et de leurs traitements). Le mot *oncologie* est également très utilisé actuellement (*oncologue*) pour « étude des tumeurs cancéreuses (évolution, traitement) ». Dans ce contexte de létalité, le terme *cancérophobie* est la peur angoissante, pas toujours justifiée, qu'éprouvent certaines personnes d'être atteintes par un cancer.

Aussi, les énoncés visent-ils à dédramatiser et le mot « tumeur » apparaît souvent associé à « bénigne » et non « maligne ». Des arguments renvoient implicitement aux politiques de prévention : « comportements à risque », « risques objectifs », « addiction ».

La présentation est prudente, pédagogique et fortement modalisée. Le terme d'origine, « crabe », persiste dans de nombreuses définitions, en lien avec le sème qui domine dans le mot « cancer », par allusion aux pinces du crabe, celui de l'animal envahisseur, trompeur, répulsif comme la maladie qui s'abat brusquement sur la personne et qui envahit les tissus et les organes.

Dans la perspective diachronique persistent ainsi trois séries de sèmes :

1. les sèmes de serrer, crabe, pinces, paires de compas, qui attrape.
2. le sème de soumois, envahissant, prolifération
3. le sème de grosseur, tumeur.

La métaphore du crabe pour le cancer reste très forte dans les textes, à cause de l'impression d'envahissement et de piège qu'elle renvoie et qui correspond au ressenti des patients. Ce terme reste encore très usité dans les discours des patients en cancérologie (Rolland-Lozachmeur 2012b).

### **3. Evolution de la définition du mot « cancer » dans le discours**

#### **3.1 L'organisation de la définition : les combinaisons syntagmatiques**

Dans notre corpus, soit 19 extraits présentés dans la partie 2, la partie définitoire est relativement courte si l'on se fie au nombre de lignes utilisées : d'une ligne à une monographie détaillée exhaustive de 23 lignes, avec des paragraphes définitoires, une délimitation, et un descriptif (au XXI<sup>e</sup> la définition de « cancer » présente les types de cancer). Les dictionnaires de langue réduisent la définition à une à 2 lignes comme Richelet. Les définitions dans les dictionnaires encyclopédiques sont étendues : Furetière alloue 12 lignes au cancer. Sur les sites Internet, la définition atteint 23 lignes pour Doctissimo, ce qui peut correspondre à une demande plus importante des usagers des sites. Les ouvrages sur le cancer présentent des textes plus développés.

Les combinaisons du mot « cancer » avec les termes génériques, les hyperonymes, les définisseurs, le groupe verbal et le groupe nominal ainsi que leurs expansions font ressortir certains traits sémantiques.

Nous traitons tout d'abord des expansions du GN et des modifications du degré. Nous observons une importante quantité d'adjectifs très variés dans l'environnement du mot cancer et de ses substituts : 224 occurrences pour 161 formes lexicales adjectivales différentes sur l'ensemble du corpus analysé. ●n relève par exemple « malignes, cancéreuses, anarchiques, cellulaire, génétique ». La plus grande partie des formes recouvre les adjectifs objectifs classifiants qui posent des ensembles stables indépendants de l'énonciateur. ●n compte ainsi de nombreux adjectifs appartenant au lexique courant ou spécifiquement médical, soit des adjectifs qualificatifs soit des adjectifs relationnels comme « ronde, inégale, gonflées, putride, squirreuses, chancreuse, anatomique, clinique, carcinomateux, cérébriforme, concrescible, hématode, cellulaire, phénotypique, oncogénique, néoplasique, a▯rabiliaire, multifactorielle, hormonaux... » et des adjectifs de couleur liés au côté descriptif « noir, cendré, plombine, livide ». Parallèlement, les adjectifs subjectifs non classifiants caractérisent et ▯raduisent en même temps la position de l'énonciateur, entre affectivité et jugement de valeur comme « hideux, insidieux, anarchique, funestes, incontrôlée, douloureuse, ingénieuse, dangereux, dissidentes, impitoyables, dévastateur, inquiétant... ».

Le verbe est le noyau de ces énoncés (Pariollaud 2008). Il se classe dans des catégories variées : verbes d'état avec une visée descriptive « être, paraître, rester » ; verbes inchoatifs pour mettre en évidence les progrès de la maladie « commencer, devenir, exister » ; verbes d'action centrés sur l'agent (le cancer, les cellules cancéreuses) « chasser, se jeter, ronger, arracher, envahir, dégénérer, dévorer, attraper, se multiplier, migrer, s'étendre, se disséminer, métastaser, permettre, proliférer, prendre pied » ; verbes axés sur le patient (le malade ou l'organe atteint) « recevoir, subir, s'écarter » ; verbes d'activité intellectuelle « savoir, prétendre, concevoir, connaître, croire, déterminer » ; verbes de sentiments « effrayer, faire peur » ; verbes de parole « dire, émettre, exprimer » ; verbes du métalangage « appeler, désigner, prendre le nom de » ; périphrases modales épistémiques « pouvoir +inf » ; périphrases actancielles factitives et passives « puisse faire connaître, faire vaciller, faire concevoir ».

Parmi les adverbes, on relève des adverbes quantitatifs du degré moyen et supérieur principalement comme « très, beaucoup, plus, fort », qui amplifient l'angoisse autour du mot ; des adverbes spatio-temporels comme « d'abord, ensuite, toujours, enfin, longtemps, tout autour, parfois » qui organisent la description et des adverbes modalisateurs comme « progressivement, constamment, effectivement, évidemment, sans doute, progressivement, anormalement, é▯rangement ».

Les relatives (69 dans le corpus) sont d'autres constantes, dans la structure syntaxique de la définition ; elles ajoutent des précisions, des caractéristiques et complètent le nom déterminé. Elles témoignent de la nécessité de développer les explications sur un sujet difficile à cerner. Ce sont des relatives adjectives essentiellement déterminatives comme « espèce de tumeur maligne qui dégénère en ulcère presque incurable » ou explicatives dans de rares cas : « terme générique désignant une tumeur maligne constituée de cellules proliférant anormalement, qui peuvent envahir les tissus voisins et donner lieu, à distance, à des métastases ».

Les termes génériques de la définition varient peu : soit le mot est pris dans un rapport autonymique et est qualifié de « terme de chirurgie, terme de médecine, terme générique », soit il reçoit des substituts comme « tumeur, espèce de tumeur, maladie, longue maladie, fléau, état pathologique ».

### 3.2 La définition dans le discours : la vérité de l'usage ?

La définition de mot représente la vérité de l'usage ; elle comporte l'idée principale qui en est la signification propre et les idées accessoires qu'elles suggèrent ordinairement (Auroux 1990, 33). Elle est ainsi à l'articulation de deux domaines (Chevalier 1990, 82) : la correspondance du langage avec la vie pour montrer comment les sens du discours acquièrent leur signification par le biais des systèmes grammaticaux et le domaine des structures linguistiques découpé en catégories qui trouve son unité dans une simplification stéréotypique.

Que ce soient les ouvrages de médecine, les dictionnaires ou les sites Internet dédiés au cancer, on retrouve les mêmes connotations de peur, de dérèglement, d'envahissement avec une accentuation sur les types de cancer et les organes touchés. On peut dire que la définition elle-même alimente la peur (« proliférer, malin, crainte, développer »). En revanche, certains mots comme « tumeur » ont tendance à être évités. Par ailleurs les descriptifs sont subjectifs par exemple dans ce passage où le lexique, et notamment les adjectifs, est subjectif ; il active le registre des émotions : « [l]e mot « cancer » choque. Il dérange. Il est cet intrus, étrangement inquiétant, invisible, sournois, impitoyable » (Paillaud, Clement-Hrynivich 2008).

La définition observe un schéma assez constant : définition (descriptif extérieur/intérieur) + actions du cancer sur la personne + cause + conséquences et évolution. Les rubriques sont déclinées par type de cancer ; ce n'est plus *une* maladie mais plusieurs types de maladie.

Il faut noter également que, dans l'usage, le terme « crabe » comme substitut de « cancer » est régulièrement utilisé dans les forums ou dans les textes autobiographiques comme un pacte de discours entre patients, ce qui

peut confirmer l'usage tabou du mot cancer souvent évité ou même peu utilisé par les médecins. On parle fréquemment dans la presse, ou dans les rubriques nécrologiques, de « longue maladie ». Parce que « cancer » renvoie au latin « crabe », cette représentation mentale menaçante a été prise comme substitut nominal, avec « écrevisse », depuis l'époque classique. Zorn (1977, 189) relève que le mot « cancer » fait peur, que sa non-prononciation est préjudiciable aux possibilités de guérison : « Personne ne se hasarde à prononcer le mot « cancer » ; rien d'étonnant à ce qu'on n'ait pas pu, jusqu'à présent vaincre le cancer. [...] Le mot est tabou. C'est ainsi que le cancéreux est condamné à désespérer totalement et à mourir de son désespoir ». Oliveira (2009, 19) montre combien métaphore et vocabulaire de la médecine sont inséparables, et que ce trope n'est pas une déviation par rapport à la norme du langage, ni un simple aspect ornemental.

## Conclusion

L'analyse des définitions en diachronie montre tout d'abord la difficulté de définir le mot et la maladie qu'il représente de manière précise, au cours des siècles ciblés par notre étude. Ce constat justifie que les reprises de certains descriptifs soient constantes au fil du temps. C'est finalement un ennemi insaisissable : une maladie qui résiste. Le deuxième trait fondamental est l'observation que les métaphores guerrières sont souvent utilisées dans les discours sur le cancer et encore à notre époque. De fait, par les implicites (présupposés et sous-entendus) et les modalisations importantes, les précautions dans le descriptif, le cancer est toujours présenté comme un mal irréductible et annonceur de mort. L'association au vocabulaire de la guerre est apparente dans certains termes comme « dépister » (c'est être sur la piste du « fauteur de trouble », « rechercher systématiquement et découvrir ce qui est peu apparent, ce qu'on dissimule, au sens de déceler, découvrir, repérer »). Ce lexique du combat pour parler du cancer est effectif dès les années 1880 (Sontag, 2009 (1977, 86). Appliquée au cancer, la métaphore occupe une place différente de celle des autres maladies comme la peste, le sida, la lèpre ou la tuberculose. Susan Sontag (*ibidem*, 66) montre qu'en comparaison de la tuberculose « l'imagerie propre au cancer est particulièrement insoutenable ». Cette remarque est à mettre en relation avec les théories sur les causes du cancer et les représentations qui en découlent. Le cancer est souvent considéré comme l'échec de l'individu à s'exprimer ; idée d'un lien entre cancer et sentiments pénibles, suivant laquelle la maladie atteint ceux qui renoncent : l'individu « perd au jeu de la vie » (*ibidem*, 68). La maladie étant vécue comme « des représailles contre l'individu qui a renoncé », il est beaucoup



question de « lutter », de « partir en croisade », « de combattre » aussi bien de la part du patient que du médecin. Le cancer est un « ennemi diabolique » (*ibidem*, 79). Le langage qui décrit le cancer évoque un événement catastrophique, l'envahissement des cellules cancéreuses qu'il faut combattre. Certes, ce langage autour du cancer, prégnant depuis très longtemps (Dupré de Lisle, 1774), évolue avec les progrès des traitements et des soins et les changements dans le langage du traitement lui-même, mais des constantes restent dans le langage courant que les dictionnaires répertorient. Néanmoins, avec le temps et l'évolution des traitements contre le cancer qui en font une maladie mieux combattue, on remarque que l'apparition du sens figuré comme emploi du mot en dehors du champ de la maladie, s'amorce : « cancer des mots » (Sartre), « cancer de l'esclavage » (Michelet) ou « cancer du plaisir » (Mauriac).

## Bibliographie

- Adelon, Bechard. 1822. *Dictionnaire de médecine*, Tome quatrième (CAB-CEV). Paris : Bechet jeune.
- Auroux, Sylvain. 1990. « La Définition et la théorie des idées », in Chaurand, Jacques & Francine Mazière, *La Définition*, Centre d'étude du lexique, CELEX, 30–40. Paris : Larousse, Langue et Langage.
- Benveniste, Emile. 1986 [1966]. *Problèmes de linguistique générale*, volume 1. Paris : Tel Gallimard.
- Branca-Rosoff, Sonia. 1988. « Les mots de parti pris. Citoyen, aristocrate et insurrection dans quelques dictionnaires (1762-1798) », *Dictionnaires des usages socio-politiques (1770-1815) fascicule 3*. Paris : INALF.
- Chaurand, Jacques & Francine Mazière. 1990, *La Définition*, Centre d'étude du lexique, CELEX. Paris : Larousse, Langue et Langage.
- Chevalier, Jean-Claude. 1990. « La place de la définition dans la Pensée et la Langue de F. Brunot (1926)- Sept remarques », in Chaurand, Jacques & Francine Mazière, *La Définition*, Centre d'étude du lexique, CELEX, 78–83. Paris : Larousse, Langue et Langage..
- Dupré de Lisle. 1774. *Le Traité sur le vice cancéreux*. Paris : Couturier, Libraire Quai des Augustins.
- Eluerd, Roland. 2000. *La Lexicologie*. Paris : Presses universitaires de France.
- Furetiere, Antoine. 1690. *Dictionnaire universel contenant generalement tous les Mots François tant vieux que modernes, les termes de toutes les Sciences et des Arts*, trois volumes. La Haye : A. et R. Leers..
- Guilhaumou, Jacques. 1989. *La Langue politique et la Révolution française*. Paris : Méridiens Klincksieck.



- oliveira, Isabelle. 2009. *Nature et fonctions de la métaphore en science L'exemple de la cardiologie*. Paris : L'Harmattan
- Paillaud, Elena & Nathalie Clément-Hrynivich, 2008. « Quand le mot cancer fait trauma. Réflexions autour de l'annonce diagnostique de cancer ». *Psycho-Onco* 2008-2, Springer, 44-48.
- Pariollaud, Fanelly. 2008. *Verbes-Questions de sémantique*. Rennes : Rivages linguistiques, Presses universitaires de Rennes.
- Polguère, Alain. 2003. *Lexicologie et sémantique lexicale*. Montréal : Les Presses de l'Université de Montréal.
- Perelman, Chaïm & Lucie ●brochts-Tyteca. 1988. *Traité de l'argumentation*. Bruxelles : Editions de l'université de Bruxelles.
- Perret, Michèle. 2008 [1998]. *Introduction à l'histoire de la langue française*. Paris : Armand Colin, Cursus.
- uemada, Bernard. 1968. *Les Dictionnaires du Français moderne (1539-1863, Etudes Lexicographiques*. Paris : Didier.
- Rey, Alain. 1990. « Polysémie du terme *définition* », in *La Définition*, Chaurand, Jacques et Francine Mazière Centre d'étude du lexique, CELEX. Paris : Larousse, Langue et Langage
- Rey, Alain. 2004 [1992]. *Le Robert Dictionnaire historique de la langue française*. Paris : Le Robert.
- Robin, Régine. 1973. *Histoire et linguistique*. Paris : Armand Colin.
- Rolland-Lozachmeur, Ghislaine, 2012a., *Approche sémantique et énonciative des discours des patients sur les forums de santé*, Annales de l'Université de Craiova, Craiova, pp.187-204
- Rolland-Lozachmeur, Ghislaine, 2012b. « Approche linguistique de la communication dans l'accès et le déroulement des soins », in *Place de la communication dans l'accès et le déroulement des soins* avec Anne-Marie Begue-Simon et Caroline HABY, *Revue Médecine*, Paris, pp.31-35
- Rolland-Lozachmeur, Ghislaine. 2015, *Le Discours sur le cancer d'un médecin au XVIII<sup>e</sup> siècle : le Traité sur le vice cancéreux par Dupré de Lisle (1774) - Analyse énonciative et lexicale*. Craiova (Roumanie) Annales de l'Université de Craiova, Sciences philologiques, Langues étrangères appliquées ANUL,X,Nr.1/2014.
- Rouëssé, Jacques. 2011. *Une Histoire du cancer du sein en ●ccident*. Paris : Springer-Verlag France,.
- Sinclair, John, Susan Jones, Robert Daley 2005. *English collocation studies, The ●sti Report- Studies in corpus and discourse*. Londres, New York : Continuum.
- Sontag, Susan. 2009 [1977]. *La maladie comme métaphore – Le sida et ses métaphores*. Paris : Christian bourgeois éditeur, collection Titres.

Tripodi, Dominique. 2009. *Facteurs pronostiques et marqueurs tumoraux de l'adénocarcinome de l'ethmoïde*. Thèse de doctorat en médecine du travail, Université de Nantes.

Wittgenstein, Ludwig. 1954. *Investigations philosophiques*. Paris : Gallimard.

Zorn, Fritz. 2011 [1977]. *Mars*. Paris : Gallimard Folio.

**PART III:**

**SFL APPROACHES TO LANGUAGE  
AND LANGUAGE CHANGE**



## CHAPTER EIGHT

### “WE SHALL SHORTLY BE ARRIVING INTO READING”:

### A CASE STUDY IN CONTEMPORARY LANGUAGE CHANGE

GORDON TUCKER

#### 1. Introduction: language change

Over the years—and certainly in recent decades—since Ferdinand de Saussure first highlighted the distinction between diachronic and synchronic linguistics, the general thrust of linguistic research has favoured synchrony. And rather than the investigation of language change over time, synchronic states of many of the world’s languages are compared and classified for similarity and difference in the field of linguistics known as linguistic typology.

Much linguistic theory sets out to account for the underlying principles of the organisation of language, principles which would essentially remain the same, irrespective of any past or present state of a given language. The organisational resources that are used by a language to express meanings may differ over time, but the range of resources itself is not typically a factor brought about by time. English, for example, has moved from being a language that exploited word morphology (e.g. noun case, verb conjugation etc.) to a greater reliance on structural resources, with minimal word morphology.

So whilst linguistic theories are concerned with the underlying principles of linguistic organisation, the data have tended to be contemporary data, and for good reason. In an electronic age, data capture has grown exponentially. We need only consider the 19 billion words of running text in the English Web 2015, web-based corpus (Jakubíček et al. 2013).

One outcome of this kind of linguistic research is description, both in terms of grammars of a given language or dictionaries, and description of current states of a language. Large comprehensive dictionaries, such as the Oxford English Dictionary (OED), provide information on historical states of entries, but regularly update entries, both in terms of contemporary usage of existing words and of neologisms.

It is not of course that the diachronic axis of language has been abandoned, but rather that those linguists who are concerned with this axis have adopted different emphases and perspectives. From the great philological tradition of charting language change, reconstructing the ancestors of our modern languages such as Proto-Indo-European, classifying languages into language families, more emphasis has been given to explaining the various kinds of language change (see Bybee 2015 and Aitchison 2013). This includes the current widespread interest in evolutionary linguistics (for an overview see McMahon and McMahon (2013). Particular aspects of language change, such as grammaticalisation (Hopper and Traugott 2003) have also received greater attention.

Within Systemic Functional Linguistics (SFL), research from a diachronic perspective has focussed primarily on functional grammatical resources that are associated with the development of scientific discourse. Whilst Michael Halliday himself opens up exploration of its development (Halliday 1988), more recently, David Banks—whose contribution to SFL and linguistics we honour in this volume—has seriously taken over the mantle (e.g. Banks 2008, 2016a, 2016b, 2019 *inter alia*). In essence, this body of work focusses on the development of scientific discourse over the centuries in terms of the use of grammatical resources such as nominalisation, thematic structure and progression, passivisation, different process types in transitivity etc., which accompany and enable the transition from instruction-based discourse, to observation, experimentation and reasoning.

## 2. Contemporary change

### 2.1 The case of ARRIVE + *into*

In this chapter I shall be concerned exclusively with language change as observed and attested over a matter of years, rather than over periods such as a century or longer, or even the period of an observer's lifetime.

Most of us, as we progress through life, and particularly as we reach “a certain age”, are aware of changes in language usage. This awareness is not different from that which concerns human behaviour in the round, whether



a matter of dress, taste in humour, music, art, for example. It is symptomatic of a natural conservatism, perhaps perfectly summed up by the refrain of the eponymous title song of the music comedy “Fings ain’t wot they used to be”, sung by the late Max Bygraves. Essentially, with new generations, conventions change, often observed with opprobrium by the older generations.

Linguist observers, again of a certain age, may well feel a degree of ambivalence in respect of this phenomenon. On the one hand, as individual users of the language, they may be subject to the conservatism mentioned above. On the other hand, from a linguist’s perspective, there is, or should be, a clear understanding of the naturalness and inevitability of language change. If one takes a millennial perspective on the issue, few speakers will extend their conservatism and nostalgia in respect of the change from Anglo-Saxon and Old English to Modern English. Instead, wherever language change is available to observation, it will be—and has been—studied as the focus of linguistic research by those who wish to chronicle it, theorise it and explain it.

Here I will attempt to crystallise this process in terms of one small personal observation, although of course observed by others, as we shall see. The observation was originally made on a railway journey from Cardiff to London, perhaps four or five years ago, and concerns a public address announcement as the train approached one of its stops, Reading. I quote, from memory:

(1) We shall shortly be arriving into Reading

My curiosity, as a grammarian, was aroused by the expression *arriving into Reading*, and in particular the use of *into* with the verb lemma *ARRIVE*. The context was also of interest. This was an official public address announcement made by the Great Western Railway train manager to passengers, most probably as part of the script prepared by this company’s relevant department, for use on all its services. The same “arrival” script was used in the approach to other stations along the route, and I have heard it used standardly on other railway journeys, even those serviced by other carriers. The utterance would therefore, at first sight, belong to a specific genre, perhaps describable as public address announcement concerning information on journey progress.

## 2.2 Expectations and priming

In terms of Hoey’s theory of Lexical Priming (Hoey 2005), the construction, in Construction Grammar terms (e.g. Goldberg 1995), *ARRIVE*

+ *into* + LOCATION runs counter to my own personal priming, where the normal expectation would be Subject + ARRIVE + *in/at* + LOCATION. I emphasise that this is a personal priming/expectation, although, as we shall see, one supported by corpus data and reference works such as the OED.

If indeed a change from ARRIVE + *in* to ARRIVE + *into* in this construction is effectively underway, an initial question concerns the locus of the change. Is the lemma ARRIVE undergoing a semantic shift or is it the preposition *into*, and how can we tell? In other words, which of the elements is responsible for the semantic coercion (e.g. Moens and Steedman 1988) if any? Semantic coercion and its relevance here is touched upon in Section 3.

In order to establish this, let us look at corpus data and dictionary entries for the two elements in the construction.

### 2.3 ARRIVE in the OED and in the corpus

The relevant sense given in the OED, considering we are concerned with ARRIVE + LOCATION is:

b. To come to the destination of a journey or reach a specified place, person, etc.; to come on the scene, make one's appearance. Later also in extended use with reference to reaching a point in a book, discussion, piece of music, etc.

The core sense here was originally used only of a person and is now the main sense of arrive.

(a) With *at, in, on, upon*, or (formerly) <sup>†</sup>*into*, <sup>†</sup>*to*, or with an adverb of place. [Thesaurus]

Interestingly, whilst the four prepositions given in the case of introducing a location are *at, in, on* and *upon*, *into* and *to* were formerly found, but now clearly marked as obsolete. The latest citation for *into* given in the OED is:

1539. C. TUNSTALL *Serm. Palme Sondaie* (1823)<sup>14</sup> Into what howse or place so euer ye shall arrive.

Typical examples from a modern corpus, here from the British National Corpus (BNC) (Aston and Burnard 1998), containing the expected range of prepositions with ARRIVE, are given in (2), (3) and (4).

(2) The Prime Minister has arrived in Rio for the final stages of the Earth Summit (BNC)

(3) Six minutes later they arrived on Planet Centram (BNC)

(4) Baldwin finally arrived at the Palace at 3.00 p.m. (BNC)

The occurrences of the range of prepositions with ARRIVE are given in Table 1. Only the prepositions introducing the destination are given, so others such as *from*, *with*, *before* etc. are excluded. It will be noted that only 3 examples of ARRIVE + *into* are found. However, 23 examples of ARRIVE + *to* are present. On inspection, however, only 8 are clearly ARRIVE + *to* + DESTINATION. And of these, only 3 are found with *to* immediately following ARRIVE. Some examples from the concordance for ARRIVE + *to* are given below in Table 2.

Table 1: numbers of relevant prepositions following the lemma ARRIVE in the BNC, from a total of 13617 occurrences of the lemma

ARRIVE + <i>at</i>	3552
ARRIVE + <i>in</i>	2414
ARRIVE + <i>on</i>	613
ARRIVE + <i>to</i>	23
ARRIVE + <i>into</i>	3

Table 2: Sample concordance lines from the BNC showing ARRIVE + *to*

I. It was then I remembered I had forgotten my boots. We	arrived	home to the sound of a ringing telephone. After a loud ex
' March into the stereo. Their guests, Michael and Helen,	arrived	to a quick apéritif of Malibu and cheese footballs. By this
ment and purpose, at the opening session. The delegate	arriving	to an indifferent reception or lack of clear information, evi
he didn't really know how to deal with his mother's grief. I	arrived	to a house stricken with grief in Plaistow Grove, Bromley,
ht of Thursday 15 June. Please note that the guests may	arrive	late to the hotel . The guests' names are: Tom Hutchinson
600 refugees from the West German embassy in Prague	arrived	to a warm welcome in Bavaria, while a train with more the
jout with truffles. Then the eagerly awaited plum pudding	arrived	to the accompaniment of clapping and cheering, and a p
rike of Canadian Racing. The Lornmores, all four of them,	arrived	together to murmurs of sympathy, but the two young one
, following numerous entanglements of rope in a tree, we	arrived	safely back to earth . I remarked it was an experience I w
from somewhere in the swathe of dry ice, The Bunny men	arrived	to polite applause and tore into a set devoid of any mater
sts first, and generally look after them from the time they	arrive	to the time they leave. It can get a bit like M15 at times - e
idence were two care workers, who said the girl regularly	arrived	back to the home in the early hours of the morning and sc
er-lit silhouettes of the Inspirals' real presence when they	arrive	to a pre-recorded keyboard and mooring groove. Althoug
rnity issued by banks to prompt the release of goods that	arrive	prior to the arrival of the original bill of lading. The reason
the Yugoslav frontier". But shortly afterwards new orders	arrived	to the effect that* No more personnel will be accepted for

### 2.3 ARRIVE and lexical aspect/aktionsart

The verb **ARRIVE** is categorised in lexical aspect terms as an ‘achievement’. Levin (2009:3) characterises this class as follows:

Some verbs, such as **explode** or **break**, describe events that are punctual; they describe the moment at which there is a transition to a result state. Thus, they are like accomplishments in being defined by a result state. Unlike accomplishments, the verb itself does not lexicalize an accompanying process, though with some verbs, such as **ARRIVE**, some unspecified accompanying process is presupposed; with others, such as **notice**, none is.

- (6) a. Carey **broke** the **window**.
- b. The **gas main exploded**.
- c. The train **ARRIVED** at the **station**.
- d. The appraiser **noticed** the flaw in the **glass**.

Such events are known as **ACHIEVEMENTS**.

The process **ARRIVE** construes motion that terminates in reaching the target location. The point at which the person/thing is at their destination is included in the semantics of the process. The dynamic/motion element explains the occurrence of the process with the present continuous tense. Unlike other achievement verbs, therefore, **ARRIVE**, as Levin points out, includes a punctual element in that “they describe the moment at which there is a transition to a result state, together with a dynamic element” (Levin 2009,3). Unlike other achievement verbs, such as **BREAK**, **EXPLODE**, **HIT**, where the duration of the event is so short as not to be registered, the duration of the event with **ARRIVE** is included. And unlike activity verbs, with durative verbs of motion, such as **RUN**, **GO**, **TRAVEL**, which have no temporal endpoint, the durative element in **ARRIVE** is circumscribed by both the temporal endpoint and by its short temporal distance from the endpoint. These observations are illustrated in (5) to (10).

- (5) the train is just arriving at platform 6 (train probably visible at the entry to the station)
- (6) ?the bomb is exploding outside the embassy (minimal duration)
- (7) we’re driving to London (said at any point of the journey)
- (8) ?we’re arriving in London (said at a good distance from London)
- (9) we’re driving (no endpoint included)
- (10) we’re arriving (endpoint included and imminent/close)

The inclusion of the endpoint/destination with **ARRIVE**, unlike with activity verbs of motion, is also indicated through the accompanying

prepositions introducing the location, primarily *at*, *in* and *on*, all of which are used with stative verbs, as shown in (11) to (13).

- (11) he lives at 36 Churchill Road (BNC)
- (12) He lives in West London (BNC)
- (13) The beach is on your doorstep (BNC)

The prepositions *to* and *into* are typically used with activity verbs of motion to indicate motion towards a location, as in (14) to (16).

- (14) he ran to the station
- (15) he ran into the station
- (16) he walked into the shop

In the model of SFL developed by Robin Fawcett and colleagues at Cardiff University, widely known now as the Cardiff Grammar (e.g. Fawcett 2000), the location element, in such examples, is analysed as a Participant Role (PR) in the structure of relational-directional clauses in English, and more specifically here as the PR of Destination, compared with the sister PRs of Source and Path (Fawcett 1987ff). This is in opposition to Halliday’s analysis of Location as Circumstance/Circumstantial Role, and in the case of these examples, the Circumstance of Location/Destination in material process clauses. Thus, the analysis of (15) here is as shown in Table 3. Furthermore, again in the Cardiff Grammar, processes such as BE and LIVE, as in *he lives in Cardiff*, are classified as Locational Relational processes, where *in Cardiff* is analysed as the PR of Location, and given the lack of directionality with ARRIVE, its Complement, as in *he arrived in Cardiff*, would also be treated as a PR of Location, rather than of Destination.

Table 3: Cardiff Grammar analysis of *he ran into the station*

<i>he</i>	<i>ran</i>	<i>into the station</i>
Affected Carrier	Process	Destination
S	M	C

S = Subject, C = Complement, M = Main Verb

When used with *on*, *in* and *at*, the prepositional Complement, as in (17) and (18), is Circumstance/Circumstantial Role (CR) of Location, rather than a PR of Destination in the process in question.

(17) he walked **at** Chatsworth (while he was at Chatsworth)

(18) he runs **in** London (he runs, as an activity, when he is within the city of London)

The behaviour of **ARRIVE**, in respect of its aspectual type categorisation is therefore supported/reinforced by its constructional association with the prepositions *at*, *in* and *on*. If we accept this as the standard/customary/stereotypical construal of **ARRIVE**, then its use with *into* suggests a “repackaging” of the form-semantics relationship of either the lexical verb or the preposition.

## 2.4 Into

The relevant sense of the preposition *into* in the **●ED**, concerning motion, is given below.

I. **●f motion ●r direction: ordinary uses.**

1.

a. Expressing **motion to a position** within a space or thing:

To a point within the limits of; to the interior of; so as to enter. In reference to a space or thing having material extension. Regularly after verbs of going, coming, bringing, putting, sending, and the like.

The essence of the sense of *into* here seems to be “to a point within the limits of; to the interior of; so as to enter”. This suggests that the resultant location is not construed as a precise location, but rather as somewhere within the interior of some space or thing, where the thing has an interior. Thus, we have the combination of *in* and *to*, where *in* signals the interior space and *to* the direction towards it. When no motion is involved, as with stative verbs, such as **WORK**, **LIVE**, **BE** etc., as in (19) and (20), *in* alone is normal<sup>1</sup>

(19) He lives **in** London

(20) He's **in** the building

---

<sup>1</sup> In some dialects of English, Bristolian in particular, *to* is found, albeit in interrogative clauses, with **BE**, as in: (1f) *Where's ee to?* (<https://www.bristolpost.co.uk/news/bristol-news/phrases-only-bristol-people-understand-7589>) (accessed December 2008)



### 3. Checking for recent change

The assumption here is that the use of *ARRIVE + into* is a recent one. This is reasonably difficult to ascertain, other than by querying corpora representative of usage over different periods of time, when such periods are in terms of years or, at most, decades. We noticed in the introduction that we were concerned here with changes over some period of the author’s life, let us say, for example, over a period of a couple of decades. Ideally, we would need corpora that were very similar in size and constitution (the text types they contain, for example), but representing respectively, say, a different decade. Given the absence of sets of corpora that satisfy these rather strict criteria, we have opted to use the BNC, the British Web Corpus 2007 (BWC07) (Ferraresi et al. 2008) and the English Web Corpus 2015 (EWC15) (Jakubíček et al. 2013), all available through The Sketch Engine (Kilgarriff et al. 2014). The BNC was completed in 1994 and contains data from the latter part of the twentieth century, as shown in Table 4.

Table 4: Distribution in time periods of texts with the BNC

	texts	w-units	%	s-units	%
<b>Unknown</b>	162	1831585	1.86	126416	2.09
<b>1960-1974</b>	46	1718449	1.74	119510	1.98
<b>1975-1984</b>	169	4730889	4.80	257962	4.28
<b>1985-1993</b>	3672	90082860	91.58	5522396	91.63

It is composed predominantly of data from the period 1985-1993 (91%) and is 89.4% written text. The general text types are as shown in Table 5. It contains overall 100 million words of British English.

Table 5: Text types within the BNC

	texts	w-units	%	s-units	%
<b>Spoken demographic</b>	153	4233955	4.30	610557	10.13
<b>Spoken context-governed</b>	755	6175896	6.27	427523	7.09
<b>Written books and periodicals</b>	2685	79238146	80.55	4395581	72.94
<b>Written-to-be-spoken</b>	35	1278618	1.29	104665	1.73
<b>Written miscellaneous</b>	421	7437168	7.56	487958	8.09

In order to provide some idea of more recent uses in English, the BWC07 and the EWC15 were used. The BWC07 was compiled in 2007 from the UK internet domain (.uk), and contains 1,313,058,436 words. The EWC15 currently stands at 19 billion words, with a target of 20 billion. It has, again, been compiled entirely from the internet. For the purposes of this study, only the .uk domain of EWC15 was selected, in terms of some kind of comparison with the British English of the BNC. However, both this 2015 sub-corpus with its 829+ million words, and the BWC07, with its 1+ billion words, are many times larger than the BNC. Consequently, this has been allowed for in the sampling size of instances found in the two corpora. Given that the lemma ARRIVE occurs 13,617 times in the BNC, a random sample of 13,617 was used from both later corpora.

Clearly, this is a crude comparison. But we should remember that what we are essentially interested in is the appearance, or increased appearance of *into*. In each case, the appearance of the prepositions in question was limited to either the first or second position after ARRIVE. This helped exclude examples in which the preposition was not introducing a location, as in *arrived for worship in ones or twos*. At the same time, however, it excludes examples where the preposition does introduce a location, but separated by another phrase as in *arrived within 10 minutes in city areas*. The number of examples for each also includes the temporal use of these prepositions, such as *arrived in the morning* and *arrived at 7.45 a.m.* Finally, metaphorical uses of ARRIVE, with a mental process sense, such as *arrived at his final opinion*, were not excluded,

Table 6: A comparison of numbers of ARRIVE + preposition across the BNC (1991-4), the BWC07 and the UK subcorpus of the EWC15), based on a sample of 13,617 occurrences of the lemma ARRIVE

	BNC 1991-4	BWC 2007	EWC-UK 2015
ARRIVE + <i>at</i>	3552	3676	3803
ARRIVE + <i>in</i>	2414	2393	
ARRIVE + <i>on</i>	613	759	722
ARRIVE + <i>into</i>	3	29	52

Despite the problematicity associated with these figures, there appears to be a noticeable pattern across all three corpora. Whilst the numbers for ARRIVE + *at/in/on* remain reasonably stable across the corpora and their respective time periods, numbers for ARRIVE + *into* have increased from 3 in the BNC to 29 in the BWC05 and to 52 in the UK domain of the EWC15.

#### 4. Semantic coercion

The apparent mismatch between **ARRIVE** and *into* suggests that one of the two lexemes is responsible for semantic coercion, in ‘imposing’ on the other some property that attaches to other elements typically associated with the other. And here, either **ARRIVE** imposes upon *into* some semantic element(s) associated with *in*, *on* and *at*, or alternatively, *into* imposes upon **ARRIVE** some semantic element(s) associated with activity verbs such as **GO**, **DRIVE**, **WALK** etc.

Given that we are dealing with a verb, **ARRIVE**, which can be categorised in terms of aspectual type, the kind of coercion at play here might itself be categorised as “aspectual coercion”. In discussions of aspectual coercion, it is usually the aspectual type of the verb that is modified in order to regularise the semantics of the whole expression.

Dölling (2014:2) gives the example in (21) to characterise the apparent mismatch of the aspectual type of the verb with the ‘adverbial’ expression.

(21) Fred played the sonata for one day

The mismatch is between the VP *Fred played the sonata* and the adverbial *for one day*, where, in the VP, the length of the playing is determined by the length of the sonata (typical performance length), yet the adverbial suggests that the sonata-playing lasted one day. The mismatch is therefore resolved by applying an “iterative” interpretation in terms of the aspectual type of **PLAY**.

In one sense, the mismatch between *arrive* and *into Reading* does seem similar to examples such as Dölling’s above, even though, of course, the preposition group *into Reading* is not a temporal Circumstance. However, there is potentially a difference in respect of language change. Despite the reader/hearer having to adjust her interpretation of *play the sonata* from activity to iterative/semelfactive, the combination of the parts would not be considered to be an example of language change. It is not that, given the combination of the two parts in (21), the verb **PLAY** will henceforth be classified as iterative/semelfactive. And yet, even with the recent usage of **ARRIVE** + *into*, is it really the case that **ARRIVE** is now to be classified as durative rather than achievement? There appears to be no evidence that, alongside **ARRIVE** + *into*, we find in contemporary usage, **ARRIVE** + *to*, as an alternative to **ARRIVE** + *at*.

If it is not the case that *into* is responsible for aspectual coercion with regard to **ARRIVE**, is it that **ARRIVE** imposes an interpretation on *into* which is semantically consonant with **ARRIVE** as an achievement verb? In other words, is *into* coerced into behaving like *in*?

In order to ascertain the direction of semantic coercion, it is arguably more fruitful to explore the occurrence of other verbs with *into*, rather than other prepositions with ARRIVE. In respect of the properties of ARRIVE, the only other preposition likely to be affected is *to*. We have already observed a small number of instances of ARRIVE + *to* in the BNC, and we can easily monitor this in the two web-based corpora that have been chosen for the purposes of comparison.

## 5. Other processes newly associated with *into*

Interestingly, a travel companion of ARRIVE + *into*, which I have again personally attested, is LAND + *into*, as in an aircraft returning to the ground after flight, or a ship reaching a port. And again, in the case of aircraft, it is normally produced by one of the pilots or cabin crew in the final announcement as it approaches the airport. The verb LAND is another verb of achievement, parallel to ARRIVE, in that it has a short duration culminating in a location, typically the city/airport/region/country where the aircraft touches down, as in (22).

(22) The small jet came in to **land** with a rush at Marco **Polo** Airport... (BNC)

From comparison of the BNC and EWC15, LAND + *into* appears to have come into use since the period covered by the BNC. Table 7 gives a concordance of some of the examples found in the English Web Corpus.

Table 7: A concordance of examples of LAND + *into* from the English Web Corpus 2015

predominantly at activity of the Dutch vessels and it is not	landed	into the UK. Red mullet are mostly caught between the m
raged around 500 t since then. However most of the bass	landed	into the UK is taken by small inshore vessels in a mixed g
'1896-1903) and the introduction of fibre-optic cables that	landed	into the ports of Dar Es Salaam and Mombasa in 2009. It
it into our neighbour's back garden this morning, sadly he	landed	into his greenhouse, and so there will be plenty of squash
ught in the deep waters off the east coast of Scotland and	landed	into Peterhead. It is a reasonably expensive product at ar
the sea is immoral and look forward to all the catch being	landed	into the future." "It is an absolute scandal that over 50% o
mplimentary on board drinks (which were unlimited!). We	landed	into New York's JFK airport where we were greeted by ou
an it comes to withdrawals, you can expect your money to	land	into your bank account within a single day. In case you we
ose! My favourite memory would be when I was about to	land	into Phoenix Airport at the start of the semester. We desc
the sea is immoral and look forward to all the catch being	landed	into the future." "It is an absolute scandal that over 50% o
mplimentary on board drinks (which were unlimited!). We	landed	into New York's JFK airport where we were greeted by ou
any other species. However, of the 12,000 tonnes of hake	landed	into the UK last year, just 1.5 per cent was consumed in th

Similar to **ARRIVE**, the three main prepositions which precede the location noun are *on*, *in* and *at*. Given the focus on achievement verbs relating to travel, another, **DOCK**, was queried, and again, was predominantly found with the prepositions *in* and *at*, as in (23).

(23) Three vessels from Tianchang, China, **docked** at the **Port of Anchorage** this week

However, a number of examples of **DOCK** + *into* were found, where the location was a harbour or port, as illustrated by (24) to (26).

(24) Every year the **location** is the same- a barge that is **docked into** the **touristic harbour of Mangalia**. (EW15)

(25) heaved a huge sigh of relief as we **docked into** Sabtang (EW15)

(26) Crashing sounds of thunder echoes through the atmosphere as we **docked into** Hanifaru Bay. (EW15)

The observations made here, arise from personal experience, such as the train manager and pilot announcement and in the context of personal primings. Any occurrences encountered in this way can subsequently be explored in appropriate corpora. Uncovering occurrences that have not been personally experienced is somewhat more complex, even with the availability of large corpora. The full EWC15, for example, contains **20,489,688** occurrences of the preposition *into*! In order to explore the full extent of the “incursion” of *into* in grammatical contexts where conventionally *in* is expected, an extraordinarily lengthy task would be required. Essentially, it would require the examination of every process, realised by a lexical verb which precedes *into* and its introduction of a Participant or Circumstantial Role. And to a considerable degree, this would require the subjective assessment of the observer, albeit an informed assessment based upon the kind of grammatical principles discussed in Section 2 above. This is therefore more of a qualitative, rather than a quantitative exercise, although once an instance has been identified, its degree of occurrence may then be quantified.

## 6. Interpretation

Any attempted interpretation of the *into* phenomenon will, arguably, be speculative. As we have observed, there is clearly evidence of a pattern of linguistic behaviour in which a group of achievement verbs involving travel **ARRIVE**, **LAND** and **DOCK**, are now frequently found with the preposition *into*, rather than the expected *in*. Let us explore, then, speculatively, some



of the possible interpretations of this change in behaviour. Reasons and explanations for types of language change in general are offered in Aitchison (2013) and Bybee (2015). Here, however, I use my own terminology for explanatory categories on phenomena affecting language change.

### 6.1 User error or misunderstanding

The seeds of permanent language change are often found in speaker error, by which we understand some misunderstanding or unawareness of an established linguistic convention. In the case here, the misunderstanding would concern the conventional use of the non-directional prepositions *in* or *at* with processes such as **ARRIVE** and **LAND**. There is then the question of whether it is misunderstanding in respect of processes such as **ARRIVE** and **LAND**, or of the preposition *into*. However, it is difficult to envisage a noticeable change in the semantics of either the processes or the preposition, a semantic change much more visible with the use of *disinterested* for *uninterested*, for example. We should also remember that the **ARRIVE/LAND** + *into* combination appears to be very much a part of official PA announcements on trains and aircraft.

### 6.2 ‘Portmanteau’ expressions, anacoluthon or semantic conflation?

Another potential explanation involves the conflation of two processes associated with motion towards a location. Examples (29) to (32) illustrate a number of processes, which unlike **ARRIVE** and **LAND**, are conventionally associated exclusively with *into*. Indeed, **DRAW** + *into* covers roughly the same short period of movement as **ARRIVE**, that is, typically within sight of the destination. For example, (29) could not be construed as taking place midway between Ghent and Bruges. This would apply also in the case of **COME** + *into*. The difference between **ARRIVE** and **DRAW/COME/FLY** + *into* is that the former lexicalises the moment of reaching the destination element included in its sense.

- (29) But the train drew into Bruges (BNC)
- (30) He saw they were drawing into Didcot Station (BNC)
- (31) It was rumoured that the ‘train’ actually came into Hallington Station (BNC)
- (32) Now 32 carriers are flying into Tegel Airport (BNC)



Clearly, at the moment of the announcement, usually within minutes of the arrival, the train is moving and the plane is flying. The preposition *into* captures this movement, whereas *in* or *at* does not. The use of **ARRIVE**, rather than **COME** or **DRAW**, lexicalises the inclusion of the arrival, in other words, and importantly for some passengers, expresses the fact that the train will be stopping at the station or physically touching down at the airport. At the same time, even though some movement is implied with **ARRIVE**, it is not highlighted, particularly in its combination with *in*, *at* and *on*. Thus, we might interpret **ARRIVE/LAND** + *into* as conflating both the element of movement with that of reaching the destination. This interpretation is shown schematically in Table 8.

Table 8: the conflation of the elements [+arrival] and [+movement] in the construction **ARRIVE** + *into*

[+arrival] [-movement]	the train	will soon be	arriving	in	Reading
[+arrival]			↓		
[+arrival] [-movement]	the train	will soon be	arriving	into	
[+movement]				↑	
[-arrival] [+movement]	the train	will soon be	drawing	into	Reading

The preference to build dynamism/movement into the semantics of a construction might also explain the wider use of *into* with processes such as **INVEST**, also personally attested, which traditionally would expect *in*.

### 6.3 Intentional semantic reduction in the difference between the prepositions *in* and *into*

To some extent this could be construed as a misunderstanding of the essential difference between *in* and *into*, although it might well be a question of users simply attributing no substantial difference between the two. With a number of processes involving movement, e.g. **GO** and **PUT**, both prepositions are commonly used. Likewise, **GET** + *in* and **GET** + *into*, with

the sense of ‘enter’, are both found in similar contexts, as shown in (32) and (33).

(33) the last thing you should do at nine o'clock in the morning is get in the car and drive the damn thing (BNC)

(34) then she went out and got into the car (BNC)

With the apparent freedom of choice between *in* and *into*, in such cases, it is not difficult to entertain the inclusion of the same freedom of choice with processes such as ARRIVE and LAND. Essentially then, the semantic difference between *in* and *into* is seen to be narrowing, essentially moving towards greater synonymy, in contexts where originally, and still for many, one or the other is expected.

## 7. Further examples

Beyond the more substantial contexts explored above, e.g. in the case of verbs such as ARRIVE, LAND etc., to the alerted observer, and here the present author, new instances surface all the time. In the course of writing this chapter, the following three examples in (35), (36) and (37) were heard during television and radio broadcasts:

(35) We've still got time into this game (BBC co-commentator, during the Cardiff versus Pontypridd traditional New Year's contest, 01/01/2019)

(36) I don't want to get involved into the politics of this (caller during LBC radio phone-in programme)

(37) Derby County are back into this cup tie (commentator, BBC TV coverage of 3<sup>rd</sup> Round FA Cup match between Southampton and Derby County, 16/01/2019)

These are, naturally, anecdotal, not part of a research-based, planned data collection. They do, however, suggest that there appears to be an increased interchangeability between the two prepositions. Furthermore, such isolated examples may subsequently be checked against corpus data. In the case of *involved into*, the UK subcorpus of the EWC15, returned 52 examples, a selection of which is given in the partial concordance in Table 9.

Table 9: concordance examples for the query *involved into* in EWC15

membrane study. These peptides were then	<b>involved into</b>	and/or weeks forming care nuclei and set for
a system. In the meantime he was actively	<b>involved into</b>	the development of a database for structural
ugh without exercising their right of getting	<b>involved into</b>	the intellectual burden of weighting "risks vere
I thank you for your generosity and getting	<b>involved into</b>	our Easter meeting. Probably the best Polish
every corner and this is your chance to get	<b>involved into</b>	something new. Last semester, you might rem
eat down by vicious babes. Slavegirls are	<b>involved into</b>	erotic power games with strict dominatrixes. M
n Napier University were I was extensively	<b>involved into</b>	social policy and policy evaluation research, a
modern world requires regions to be actively	<b>involved into</b>	global integrating processes, broaden econor
r and next winter? Marc: I'm looking to get	<b>involved into</b>	the professional training program at our ski sc

Amongst the 65 concordance lines returned for the *involved into* query, 13 were not genuine examples of the phenomenon, since *into* related not directly to *involved*, but to the wider clause, that is, *forcing someone into something* as shown in (38).

(38) not to force tenants involved into a long drawn out legal battle for fair play (EWC15)

## 8. Postscript

The full implications of this observed change in the use of *into* may not be understood until much later. Not all short-term language change is successful; for example, lexical neologisms or re-semanticisation of lexical items at the hands (or mouths) of subgroups in society do not always acquire permanency. Indeed, they often manifest the same tendencies that are found in sartorial fashion, hairstyles, modern music, and fall into disuse, particularly by those, typically the younger generations, who introduced them in the first place. One observed consequence of this is that their later acquisition by others, particularly older generations, takes place after they have fallen into disuse, giving rise to hilarity or disdain on the part of their original creators in respect of their older “misusers”.

Currently, there would appear to be an increase in the use of *into*, where once *in* was expected. Only time will tell whether this tendency continues and establishes itself firmly and canonically in the language.

From an SFL perspective, in theoretical and descriptive terms, there is little that can be done with such phenomena, other than observe and

describe. Ultimately, with further evidence of the tendency and its various contexts, the new relationship between, processes such as ARRIVE and LAND, and the preposition which introduces their corresponding Participant Role of Location, will need to be registered in the appropriate system network and realisation rules. But this is a problem for all linguistic description. Languages, unlike the artist's expert model, refuse to stand still; they are constantly on the move!

## References

- Aitchison, J. 2013. *Language Change: Progress or Decay*. (4<sup>th</sup> Edition). Cambridge: Cambridge University Press.
- Aston, G. and L. Burnard. 1998. *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.
- Banks, D. 2008. *The development of scientific writing. Linguistics features and historical context*. London: Equinox.
- Banks, D. 2016a. "Diachronic aspects of ESP". *ASp* [Online], 69. URL: <http://asp.revues.org/4812> ; DOI : 10.4000/asp.4812.
- Banks, D. 2016b. *The Birth of the Academic Article: Le Journal des Sçavans and the Philosophical Transactions, 1665-1700*. Sheffield: Equinox.
- Banks, D. 2019. "SFL and diachronic studies". In G. Thompson, W. Bowcher, L. Fontaine and D. Schönthal (Eds.) *The Cambridge Handbook of Systemic Functional Linguistics*. Cambridge: Cambridge University Press.
- Bybee, J. 2015. *Language Change*. Cambridge: Cambridge University Press.
- Dolling J. 2014. "Aspectual Coercion and Eventuality Structure". In Robering, K. (Ed.) (2014): *Aspects, Phases, and Arguments: Topics in the Semantics of Verbs*. Amsterdam: John Benjamins, 189-226.
- Fawcett, R.P. 1987. "The semantics of clause and verb for relational processes in English". In *New Developments in Systemic Linguistics, Vol 1: Theory and Description*, M.A.K. Halliday and R.P. Fawcett (Eds.), 130-183. London: Frances Pinter.
- Fawcett, R.P. 2000. *A Theory of Syntax for Systemic Functional Linguistics. Current Issues in Linguistic Theory 206*. Amsterdam: John Benjamins.
- Ferraresi, A., E. Zenchetta, M. Baroni and S. Bernardini. 2008. Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the 4th Web as Corpus Workshop (WAC-4) Can we beat Google*. 2008, pp. 47-54.

- Goldberg, A. 1995. *Constructions. A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
- Halliday, M.A.K. 1988. “On the language of physical science”. In M. Ghadessy. (Ed.), *Registers of Written English: Situational factors and linguistic features*. Pinter: London, 162–178.
- Hoey, M. 2005. *Lexical priming: a new theory of words and language*. London and New York: Routledge.
- Hopper, P.J. and E.C. Traugott. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Jakubiček, M., A. Kilgarriff, V. Kovář, P. Rychlý and V. Suchomel. 2013. “The TenTen corpus family”. In *7th International Corpus Linguistics Conference CL* (pp. 125-127).
- Kilgarriff, A., V. Baisa, J. Bušta, M. Jakubiček, V. Kovář, F. Michelfeit, P. Rychlý and V. Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography*, 1: 7-36.
- Levin, B. 2009. *Lexical Semantics of Verbs IV: Aspectual Approaches to Lexical Semantic Representation*. Course Notes, UC Berkeley and Stanford University. <https://web.stanford.edu/~bclevin/lisa09aspapp.pdf> (Accessed 20.12.2008).
- McMahon, A. and R. McMahon. 2013. *Evolutionary Linguistics*. Cambridge: Cambridge University Press.
- Moens, M. and M. Steedman. 1988. “Temporal ontology and temporal reference”. *Computational linguistics* 14(2), 15–28.
- OED Online. 2013. Oxford University Press. <http://www.oed.com.abc.cardiff.ac.uk> (accessed December 2018).





## CHAPTER NINE

# WHEN LEAVING IS A THING: CONSTRUING BRITAIN'S EXIT FROM THE EUROPEAN UNION

LISE FONTAINE

### 1. Introduction

Few would disagree that words such as *leave*, *exit*, *withdraw*, etc. are verbs which encode an event, a happening, a process. For Langacker, verbs such as these are best considered in terms of construal, i.e. how they are processed “sequentially through conceived time” (1987a:55). Halliday holds a similar view of verbs, which realize or construe a process “as being located in, and unfolding through, time” (Halliday & Matthiessen, 2014:212). The verbs mentioned above also have nominal counterparts in *leaving*, *exit*, *withdrawal*. The case of *exit* is interesting because its status as a noun or a verb is morphologically unmarked, unlike ‘leave/leaving’ and ‘withdraw/withdrawal’. These nominal forms differ both semantically and instantially, from their verbal counterparts. In many cases nominalization does maintain a high degree of semantic overlap with the root verb but as is generally accepted in the literature (e.g. Halliday, 2006, Langacker 1987a), each involves a different construal.

According to Banks (2008:130), “[I]t is evident that the use of nominalisation is rooted in the situation. The nature of the activity going on to some extent determines the form of the language used”. *Brexit*, a neologism coined in 2012, evolved in a particular situation, one that has been reasonably well documented through written digital media. It is unusual to find a new word entering the language in such an identifiable and recorded way as was the case for *Brexit*. The term itself is used to refer to the event of the United Kingdom of Great Britain and Northern Ireland (UK) leaving the European Union (EU). Therefore, as a noun, it is clearly

denoting a process rather than an object but clearly the process semantics have remained.

Fontaine (2017b) presented the argument that the term *Brexit* could be considered as a complex nominal. The main reason given for this was the incongruence of using *Brexit*, a noun, to express a process. In other words, by Halliday's account of grammatical metaphor (GM), the process meaning of 'exit' in *Brexit*, has been reworded metaphorically as a noun rather than congruently as a verb. Evidence for this position was based on corpus data which supported this interpretation of *Brexit* since the term was typically marked, in its earliest uses, as novel, providing a gloss, typically in apposition, such as 'Brexit, short for British exit', 'Brexit, Britain's exit from the European Union' or 'Brexit, Britain leaving the EU'. However, another conclusion drawn was that "we cannot assume that the relationship necessarily involves congruence" (Fontaine, 2017b:14). This chapter takes up this point in relation to congruence and seeks to resolve whether or not we can find any evidence that congruence, or incongruence, is a determining feature of the lexical item *Brexit*. While incongruence is a motivating concept for grammatical metaphor (GM), in this chapter we will also consider *Brexit* from the perspective of construal, a concept which, within SFL, is very closely aligned with congruence.

The chapter is organised as follows. A short history of the lexeme *Brexit* is presented in order to establish the context in which *Brexit* has developed. Following this, the concepts of congruence and construal are examined in relation to the lexeme *Brexit* in terms of how they are defined and used within SFL. In section 4, the construal of *Brexit* is explored through its collocations in two time periods, 2012-2015 and 2015-2017. The chapter concludes by highlighting the difficulty in establishing congruence for complex nominals such as *Brexit*, and by suggesting that construal offers a potential source on which to base the development of empirical evidence for congruence.

## 2. A short history of the lexeme Brexit

*Brexit*, as a neologism, was first coined in 2012. To the best of my knowledge, it appeared in a blog in May, 2012, followed by a mention on Twitter, as shown in example (1). I was not able to find any earlier recorded instances.

- (1) British Influence @britinfluence 15 May 2012 Stumbling towards the Brexit - Britain, a referendum and an ever-closer reckoning: Nucleus News <http://eepurl.com/lRyND>

As will be explained in this section, two significant events in the UK have propelled the word into common use and by the time of the referendum on EU membership held on June 23 2016, *Brexit* was a household word and at the time of writing, January 2019, it would be fair to say that its frequency of use has increased exponentially. In the short span of time in terms of a word's history, *Brexit* is no longer novel and is readily taken for granted as an established term for reasons that will be discussed below. *Brexit* is a blend of *Britain* or *British* with *exit*. Its formation was influenced by analogy to the term *Grexit*, which is also a relatively recent blend (*Greek* + *exit*), although older, and which was coined in response to the Greek debt crisis in 2010. Both *Grexit* and *Brexit* capture the meaning of 'a country's exit from the European Union'. The relationship between *Grexit* and *Brexit* was quite salient around the time that *Brexit* was coined, as highlighted in example (2)<sup>1</sup>, which includes incidentally the term 'QEExit', which has clearly not survived.

- (2) The original term 'Grexit', coined to describe the fear that Greece might leave the euro, has now inspired the neologisms 'Brexit' (concern about Britain quitting the EU) and 'QEExit' - the threat that the Federal Reserve (the Fed) will cut back on QE. (source: Money Marketing, July 12, 2013)

In order to build a picture of the word itself and how it was used in the first few years of its development, two data sets were compiled. The first covered the period May 2012 to May 2015, as described in Fontaine (2017b). A corpus of 1,641,903 tokens was constructed with a frequency for *Brexit* of 2,435 instances. The second covered the period May 2015 to June 2017 with 11,075 instances of *Brexit* in a corpus of 3,913,169 tokens. May 2015 and June 2017 represent significant dates in the UK since at these times a general election was held where membership of the EU was a significant issue. The texts were gathered using Lexis Nexis Academic, an online database that contains full text newspaper articles from around the world in a variety of languages<sup>2</sup>. The search term used was 'Brexit', which resulted in a corpus of English language news articles which included at least one mention of this term. Due to the volume of articles mentioning

<sup>1</sup> Unless otherwise indicated all examples are taken from the online database Lexis Nexis Academic as described in this paper.

<sup>2</sup> Data used in this paper was obtained from publicly available newspaper articles, which were accessed via the LexisNexis Academic website: <http://www.lexisnexis.com/hottopics/lnacademic/>. Accessed 7 December 2016. The texts included will not be shared due to the Terms & Conditions of use for the LexisNexis Services.

Brexit, the corpus constructed for this study was not exhaustive, i.e. not every article in this time period was included, but rather a sufficient number of articles was included over the period to ensure the corpus size was large enough to obtain collocates. In Table 1 we can see that the normalised frequency per million words in the second data set (2015-2017) has more than doubled (an increase of 1.58%). It would be reasonable to predict that in the 2017-2019 period, this frequency would have continued to increase exponentially.

Table 1 Frequency of Brexit in 2012-2015 and 2015-2017

	2012-2015	2015-2017
Instances of BREXIT	2,435	11,075
Corpus size (# words)	1,641,903	3,913,169
/ million	1,483.04	2,830.19

Consequently, the full data set includes a substantial number of instances of *Brexit* over the first five years of its use, from the time that the term was first coined in May 2012 to the UK election on June 8 2017. SketchEngine (Kilgarriff et al, 2014) was used to analyse the corpus in terms of concordances and collocations and to extract examples from the corpus. While non-English examples were excluded from this corpus, it is clear from examples (3) to (5) that even in 2012, before it was a common word in the UK, the term had gained some currency in German, Spanish and French language newspapers.

- (3) *Börsen-Zeitung* Freitag 2. November 2012  
Was niemand braucht, ist eine weitere Abkopplung Großbritanniens von der EU. Nicht zufällig hat der Albtraum der EU-Spitzen mittlerweile mehrere Namen: Nicht bloß Grexit, sondern auch **Brexit**.
- (4) *El País* October 26, 2012 Friday  
Primero fue Grexit, la combinación de Greece (Grecia) y exit (salida). Ahora es **Brexit** o **Bríxit** para referirse a una posible salida de Reino Unido de la UE (British exit).
- (5) *La Tribune* Lundi 22 Octobre 2012  
C'est le nouveau néologisme à la mode dans les milieux financiers et politiques, « **Brexit** ». Une salade linguistique assez surprenante issue de « Grexit », lui-même désagréable mélange de « Greece » (Grèce) et « Exit » (sortie), cette fois agrémentée du B initial de « Britain », Grande-Bretagne.

●n tremble de voir arriver sur nos écrans et nos journaux les futurs « Spexit », « Itexit » ou, pire, « Frexit », avant le niveau ultime de la crainte universelle, « Gexit » ...

These examples demonstrate how politically charged the term *Brexit* is and how it is clearly based in economic and political discussions of the UK's membership and relationship to the European Union.

In the early uses of *Brexit*, we do find spelling variations such as *Brixit*, but this instability did not last, and *Brexit* was quickly established as the only form of the lexeme. Despite it being clearly a blend of two words, recovering the 'br-' part of the blend is not necessarily obvious. It could be the onset of *Britain* or *British*, e.g. *a British exit from the European Union* or *Britain exits the European Union* or *Britain's exit from the European Union*. In some respects, the source of the BR- in *Brexit* is not significant, since either form identifies who or what is engaged in the exit and this is overtly marked. However, it is worth pointing out that it is in fact the government of the United Kingdom of Great Britain and Northern Ireland, ultimately the UK Prime Minister, as the agent of exiting the EU. Nevertheless, it is reasonable to analyse *Brexit* as a blend of *Britain* or *British* and *exit*. According to the Oxford English Dictionary (OED), 'exit' is a borrowing from Latin as a noun, from some time prior to 1564, and has rather complex origins. It entered the English language as a noun of action from two formations ("Exit" v. and n.): (1) < Latin *exit*, 3rd person singular indicative of Latin *exīre* 'to go out', < *ex-* 'out' + *īre* 'to go' and (2) < Latin *exitus* (u- stem) 'going out, departure'. The English verb EXIT (meaning 'to depart') was formed in English by conversion from the noun (i.e. *exit* v. < *exit* n) ("Exit" v.) and is attested from 1607. Therefore, in the source language, Latin, 'exit' clearly had grammatical status as a verb but in English it takes up its place in the language as a kind of verbal noun, a noun of action, with the potential to express a process meaning.

Verbs are readily created from nouns and vice versa. This process of flipping a word from one word class to another, without any morphological marking is referred to as conversion (or zero derivation). Conversion is a very productive and common word formation process in English. Barber (1997:238) explains that in the Middle English (ME) period, the most common type of conversion was denominal verbs, i.e. verbs derived from nouns. Given that the formation of *exit* in English takes place in the ME period, it is reasonable to assume that its conversion would have been supported by the same process happening to other nouns at the time.

The formation of *Brexit*, however, was not due to conversion but rather to a blend, even though *exit* as part of the blend carries in it aspects of the verb conversion. As Plag (2003:155) explains, a blend "involves two or

(rarely) more base words (instead of only one), but shares with truncations a massive loss of phonetic (or orthographic) material". Examples of blends in English include *brunch*, *blog* and *emoticon*. Blends not only blend sounds but also meanings and behind them the conceptual meanings of the words blended together. According to Benczes (2006: 58), "[t]he language user needs to construct blends on the basis of conceptual metaphors and metonymies and pre-existing frames of understanding, which are then moulded in the blend to fit the discourse context". This is an important aspect of blends that we will return to below since it contributes to the semantic complexity of the term.

In the early examples of *Brexit* as shown in (6) to (9), we find support that people using the term *Brexit* saw it as a blend of *British* and *exit*.

- (6) THERE'S a new word on the lips of Eurocrats at the moment: **Brexit**, short for **British exit**. (Express, 26 Oct 2012)
- (7) Mr Cameron's much-trailed Europe speech, expected in mid-January, will seek to tread a delicate line, promising renegotiation to avoid a **Brexit**, or **British exit** from the EU. (The Telegraph, 28 Dec 2012)
- (8) The European media had been filled for weeks with speculation about a **Brexit**, a **British withdrawal** from the European Union, partly pushed from Europe and partly pulled by the **British**. (UPI, 26 Nov 2012)
- (9) The so-called **Brexit question** - the issue of whether **Britain** is heading for the EU **exit door** - has been a few years in gestation in **Brussels** (The Guardian, 27 Dec 2012)

What we notice in these examples is an awareness of how unfamiliar and new *Brexit* was likely to be to the readers and as the term was used, this novelty was signalled, often with a gloss, quotes or expressions such as 'so-called', as highlighted in bold in the examples. Bauer (1983:42) explains that:

When a word first appears in a language, whether as a loan or calque, or as a nonce formation, it appears that speakers are aware of its newness, that is they are aware that they are exploiting the productivity of the language system. Thus, in modern journalistic language the word is often put in inverted commas, a phrase is added such as "what has been called", "as it is termed" and so on, or a complete gloss is provided.

Early uses of *Brexit* indicate explicitly a recognition that the productivity of the language system is being exploited through the use of 'so-called' and scare quotes, and explanations and definitions of the term in apposition or in a non-restrictive relative clause (cf. Goatly's (1997, 171-175) typology of "definitional contexts"). However, neologisms such as this do not always



survive and the lifespan or “currency” (Bauer, 1983: 43) of a word is affected by many factors, which include, according to Bauer: the status of the person using the term, the attitude to the word evinced by society as a whole, and whether or not there is a need for the form.

As we know now, *Brexit* was a neologism with a great deal of currency and it has certainly survived. The 2015 general election result gave *Brexit* a significant boost because the Conservative Party promised a referendum on EU membership. *Brexit* then became a very convenient and efficient term to capture, or perhaps mask, the complex question surrounding the UK’s membership in the EU.

While not appearing in dictionaries until 2016, it first appeared as a term on Wikipedia in January 2014 and as an entry in 2016. *Brexit* was named the word of the year 2016 by Collins Dictionary. According to Bains (2016), “Lexicographers recorded an upsurge of over 3,400 per cent in the use of ‘Brexit’ this year – a level of increase unheard of since Collins began monitoring word usage.”

At the time of writing (January 2019), talk of *Brexit* is certainly dominating in the media, online, and, it seems, everywhere you turn. The UK has legally established its intention to leave the EU and has set a date of March 29 2019 when it is scheduled to withdraw from the EU. It will certainly be very interesting to see what happens to this lexeme if the UK really does leave the UK. It is difficult to imagine that it will cease to be used. As we will see below, we have already seen a shift in its grammatical function from that of Thing (or head of the noun phrase) to having a role as a modifier (Classifier in Halliday & Matthiessen’s (2014) terms). One might predict that if the UK does leave the EU, it would likely become more of a proper name, used as a name to refer to the event as a uniquely identifiable entity, and no longer modifiable, perhaps to some extent similar to national events such as Independence Day in the USA.

### 3. Congruence and construal

The terms congruence and construal, like many terms in linguistics have general uses but within linguistics there are differences in how they are used technically. Congruence has been used to discuss processing language (e.g. Clark’s (1974) principle of congruence) and for cross-linguistic comparison (e.g. Ebeling, 2018). Further congruence is often equated with agreement (e.g. gender agreement between adjectives and nouns in some languages). In SFL, congruence is used to describe a type of relation between lexicogrammar and semantics, which are represented as two different strata in the model. The idea that underpins this concept is that there is a

congruent, or typical, relationship between the units of the lexicogrammar (e.g. nominal group, clause) and the semantics. According to Hansen-Schirra, Neumann & Steiner (2012:135) “‘Congruent’ refers to a direct and transparent mapping from semantic onto grammatical categories, for example processes onto verbs, entities onto nouns, qualities of an entity to an adjective, etc., whereas other ‘indirect’ mappings [...] are classified as ‘metaphorical.’” For example, as concerns mood, Halliday & Matthiessen (2014:195) claim that “[f]or statements and questions there is a clear pattern of congruence: typically, a statement is realized as declarative and a question as interrogative.” However, once the concept is confronted with empirical data, it is not so clear what relations should be treated as congruent and which should not. As explained by Steiner (2004:144), “notions of ‘directness’ or ‘congruence’ [are often introduced] as pre-theoretical givens into the discussion [lacking] theoretical motivation and clarification”.

As put forward in Fontaine (2017b), the status of the use of *Brexit* in terms of congruence is not unproblematic. If we accept the assumption that process semantics (event meanings) are congruently realised by verbs and thing semantics (object meanings) are congruently realised by nouns, then in *Brexit*, we find a noun that expresses process semantics. However, it seems clear from the data that it was originally conceived as nominal and that even among the explicit glosses of the term in the early data presented in Fontaine (2017b), we typically find nominal, not clausal, expressions in apposition, e.g. *a British exit, a British withdrawal*. Readers will certainly infer that the semantics include *exiting* and *withdrawing*, but this event has not yet occurred and while its hypothetical status is not part of its congruence, it is part of its construal as we will see below. If we accept that the realization is incongruent, we must ask what the congruent realization would be. It becomes, with this line of thinking, somewhat more difficult to think of *Brexit* as a metaphorical variant of some other available congruent encoding.

Construal is closely related to congruence. It is a theoretical term with two main uses within SFL (Matthiessen, Teruya, and Lam 2010:76): to create meaning in the ideational mode, i.e. construing experience; and to refer to realization as in “grammatical patterns construe semantic patterns”. It is construal then that gives meaning to the instances of language. For example, an imperative clause such as ‘eat your vegetables’, is said to construe a command. It is the process of meaning-making; we might think of it as a kind of inference, or interpretation. As Taverniers (2011:1122) explains, construal refers to “the relationship between language and extralinguistic reality”. In relation to congruence, as discussed above, construal is generally used to explain the relationship between semantics

and lexicogrammar (cf. realization) in order to specify whether the relation is construed congruently or not. Therefore, anything expressed by a nominal group would, in theory, be being construed as a thing. However, we must ask whether we are saying the same thing if we say that a 'thing' is realized by or is construed by a nominal group. It is certainly clearer if we use realization in relation to congruence/incongruence and leave construal as a separate concept, albeit very much intertwined with both congruence and realization. Wegener (2011:5) sees realization as having a dialectal relationship with construal and construction. Hasan (2010:12) explains this relationship as follows: "In the reception of the utterance, realization is construal of the relevant choice at the higher level: thus in decoding an utterance, the choice in wording construes meaning, the choice in meaning construes context." Here we can see that realization and construal are nearly equated in terms of, in the one sense, relating the wording to the semantics but we also find simultaneously that the wording construes experience, i.e. is meaning-making in ideational terms.

A congruent construal is based on an assumption of a typical relation between the wording and the semantics; as a class, a noun construes a thing, i.e. nominal wording (lexicogrammar) typically realizes a participant (semantics); where this is not the case, e.g. where a nominal construes a process, the construal would be said to be incongruent. For Halliday, this is where the concept of grammatical metaphor comes in:

Grammatical metaphor is what turns *move* into *motion*, *resist* into *resistance*, *fail* into *failure*, *long* into *length*, *can* into *possible*, *so* ('therefore') into *cause* (verb or noun). It is metaphor because it involves cross-coupling between semantics and lexicogrammar: an expression is being used to mean something that has usually been meant by something else. (Better: **a meaning that has usually been realised in one way is now being realised in another**.) It is grammatical because what is being cross-coupled is not a word (i.e. not a lexical item or 'lexeme') but a class: a noun is doing the job of a verb or adjective, an adjective that of a modal verb, a verb is doing the job that has been done by a conjunction. (Halliday 2006:114, emphasis added)

The difficulty we face with a lexeme like *Brexit* is that it expresses a meaning that has only been realised in one way, as a nominal and as we saw above, it is formed by analogy to *Grexit*. While it is morphologically a blend, it also seems to be a blend in Halliday's (1994:399) sense: "Blends construe two (or more) different meanings; but the meanings are fused - it is not a matter of selecting one or the other." This provides an alternative perspective on construal. Wording can construe more than one meaning. In

the case of *Brexit*, the meanings of thing and process are simultaneously construed and yet we might feel uncomfortable accounting for this in terms of an incongruent realization.

Given that *exit* (at least in the sense evoked in *Brexit*) can be sudden or prolonged (e.g. 'His exit is lasting hours) and it can be observed (e.g. 'I watched his abrupt exit'), we can say that it behaves, aspectually, as a process. Indeed, many in the UK will be thinking that *Brexit* is lasting years. The term is, however, actualised in wording very clearly as a nominal. I have argued in the past that not all nominalizations involve the same degree of complexity and also not all the devices for deriving nominals involve the same degree of complexity. However, what does seem to be apparent is that English has developed a very productive means of converting nouns into verbs and vice versa (see Fontaine, 2017a). Halliday (1990/2007: 363) sees this fluidity between nouns and verbs as an instance of complementarity, a particular type of indeterminacy where we can "[interpret] what goes on either as a construction of objects or as a flow of events". As an explicit blend, *Brexit* appears to allow both.

If we consider *Brexit* as an incongruent realization, i.e. as grammatical metaphor, we can only do so if we assume incongruence because a process has been reworded metaphorically as a noun (Halliday & Matthiessen, 2014:729). However, an extreme interpretation of this, as argued in Fontaine (2017a), could lead to treating event-nouns such as *storm* as an instance of grammatical metaphor; after all, if the wording construes an event, but does so incongruently, would this not be the case? Part of the complexity is that in terms of the morphology of *Brexit*, its composition as a blend includes the construal of agency in that Britain will do the exiting, although this could also be considered an instance of referential metonymy since the use of *Britain/British* is to refer to the entire UK. The agency very clearly suggests a figure which is congruently realised by a clause, as illustrated in example (10), where the nominal group in apposition to 'Brexit' includes, as a qualifier, an expression of motion, *from the European Union*, which encodes a directional meaning (cf. Langacker, 1987b: 174).

- (10) More fundamentally, the prospects of a **Brexit, a British withdrawal from the European Union**, are becoming serious.

Given the complexity of the semantics related theoretically to the blend *Brexit*, the next section takes a more empirical approach and examines the top collocates of *Brexit* in each corpus in order to explore its semantics through construal, including whether there have been any shifts in use in the period following the 2015 general election which might have altered its construal.

#### 4. *Brexit* construed through collocations

As described above, two corpora were used to examine uses of *Brexit*. Both used the same source, Lexis Nexis Academic, and the same text type (English news articles) but with texts taken from two different periods: 2012-2015 represents the period from the first recorded use of *Brexit* to the UK general election in 2015 and 2015-2017 represents the period following the 2015 election to the UK general election in 2017. It is clear from the discussion given above that the frequency of *Brexit* increased dramatically in the second period, something which is also attested by Collins (Bains, 2016). In order to compare the uses of *Brexit* in the two periods, the top 10 collocations were taken from each corpus using SketchEngine (Kilgarriff et al., 2014). The collocations were calculated using a -2/+2 word span (i.e. the calculation of collocates includes two words to the left and two words to the right of the key word in context). The results are shown in Table 2 and Table 3 respectively.

In Table 2, as was also the case in the results given in Fontaine (2017b), we find collocates that evidence the explicit awareness of the novelty and unfamiliarity of the word *Brexit*: ‘so-called’ and ‘or’, as given in examples (11) and (12).

- (11) But a **so-called "Brexit"** does not seem to trouble some Europeans  
 (12) A **Brexit, or a British exit**, is becoming a serious prospect, even though losing Britain in order to keep Greece looks a very strange bargain.

The most noticeable development in the 2015-2017 collocations, as noted in Table 3, is the shift we see in the use of *Brexit* from Thing in the nominal group, as semantic head, to functioning as a pre-modifier expressing the Classifier element in a nominal compound, as in ‘The *Brexit* vote’, ‘These *Brexit* negotiations’, and ‘The *Brexit* campaign’. This distinction is illustrated in examples (13) and (14), where we can see that one effect of this is that nominal groups with *Brexit* as a Classifier can be made definite since the semantic core of the referring expression is being used not to refer to the event of the UK leaving the EU but to various aspects related to it (e.g. *vote*, *campaign*, *campaigner*, *negotiation*). There is only one dominant compound in the 2012-2015 period, and this is related to a prize, which was established in 2013 by the Institute of Economic Affairs (IEA) and which was called the ‘*BrexitPrize*’. According to the IEA website (<http://iea.org.uk>), the competition offered a prize of “100,000 euros to the individual or team who could develop the best blueprint for Britain’s future in the event of an EU referendum resulting in an ‘Out’ vote”. An example

of this is shown in example (15) below. The **Brexit prize** was awarded in 2014.

- (13) **A Brexit**, however, would not help this situation. (2012-2015)
- (14) **The Brexit vote** has made the **British** commercial environment much more uncertain and **French trade** unions, who want the **final investment decision** postponed, have been pressuring **independent directors** to convince EDF's chief executive, Jean-Bernard Levy, to **ditch Hinkley**. (2015-2017)
- (15) The organisers of the new **Brexit prize** may be guilty of a little hyperbole when they say it would be comparable to the fall of the **Berlin Wall** and the subsequent collapse of **Soviet communism** but they are not wrong to emphasise its importance. (2012-2015)

The only collocate that is shared in the two periods is the modal auxiliary 'would'. In the 2012-2015 period, we find 'Brexit would be', 'Brexit would lead', 'Brexit would cost' and 'Brexit would mean'. With 'be' being the most frequent and top collocate for 'Brexit would be' phrase, we find that the Attributes following 'be' most frequently express a negative semantic prosody, as shown in example (16). Further following 'lead', 'cost' and 'mean', the top three collocates after 'be', we find a trend toward similarly negative semantic associations as shown in examples (17) to (19).

- (16) Some make economic arguments against it too, but in fact **Brexit would be a financial catastrophe**, costing us as much as 14% of GDP, if studies out this week are to be believed.
- (17) More worryingly still, some 50 per cent of investors believe **Brexit would lead** directly to a fall in property values.
- (18) In a best-case scenario, in which **Britain** has a **trade** agreement with the EU similar to that of Switzerland, **Brexit would cost** every UK citizen the equivalent of 220 euros (£157) a year at current prices by 2030, said the thinktank.
- (19) **But** the conclusion that **Brexit would mean** a net loss for the UK is one very much shared by **Britain's business lobby** groups.



Table 2 Top 10 collocates for 'Brexit' (2012-2015)

Corpus 2012-2015	co-occurrences	candidates	T-score	MI	LogDice
so-called	132	209	11.46284	8.77186	10.72021
Prize	57	82	7.53414	8.91016	9.582
a	721	31324	25.16588	3.9937	9.45429
would	204	7300	13.54437	4.27357	9.43536
or	101	3401	9.56091	4.36129	9.16736
possible	43	529	6.44088	5.81397	8.93241
could	77	2904	8.29679	4.19779	8.90621
risk	47	1023	6.64005	4.99082	8.83264
potential	36	452	5.89115	5.78457	8.71511
fear	32	512	5.52608	5.43483	8.51467

In the 2015-2017 period we no longer find an association with 'be' following 'Brexit would' but we do find that the collocates 'mean', 'cost', and 'cause' are the top three. An example of each is given respectively in (20), (21) and (22).

- (20) Before the EU vote, David Cameron said **Brexit would mean** the end of the triplelock that ensures pensioner incomes grow with the cost of living.
- (21) He "warned that **Brexit would cost** the average household about £4,000 per annum in the future".
- (22) The lobby group recently published an analysis claiming **Brexit would cause** a serious shock to the UK economy, which could lead to 950,000 job losses and leave the average household £3,700 worse off by 2020.

Table 3 Top 10 collocates of 'Brexit' (2015-2017)

Corpus 2015-2017	co-occurrences	candidates	T-score	MI	LogDice
vote	856	9511	28.49957	5.27064	10.42081
after	511	7976	21.78269	4.78029	9.78902
mean	367	2736	18.82427	5.84634	9.77912
negotiation	316	1188	17.62058	6.83402	9.73642
would	577	16087	22.45943	3.94338	9.44973
campaigner	237	881	15.26138	6.8503	9.35834
for	895	37180	27.01905	3.36806	9.25107
back	289	5005	16.31359	4.63033	9.21311
could	291	6964	16.10694	4.16373	9.05599
campaign	263	5752	15.39035	4.29363	9.0111

Although this discussion is far from exhausting the data, there is evidence to show that there have already been some semantic shifts in the use of *Brexit* both in terms of the other lexemes with which it frequently co-occurs as well as in terms of its functional role within the nominal group. In the 2012-2015 period, the use of *Brexit* construes a sense of novelty, something new, but also a hypothetical event. While there is no longer a construal of newness with the use of *Brexit* in the 2015-2017 period, one constant was identified to hold over the first five years of the lexeme *Brexit*, and this is its strong collocation with the modal auxiliary 'would' which highlights the modality and hypotheticality of *Brexit*. It was also noted that along with the co-occurrence of 'would' we find verbs that introduce negative semantic prosody, which colours *Brexit*, at least in the data observed here, in a negative light.

## 5. Concluding remarks

This chapter has built on the initial work on *Brexit* as set out in Fontaine (2017b). This has been done by extending the time period included in its development so that the period between the UK general election in 2015 and the referendum on EU membership in 2017 could be included for comparative purposes. *Brexit* is a blend within which a variety of complex meanings and complex perspectives converge. Its complex semantics was

then considered from the perspectives of congruence and construal. I suggest that while the concept of (in)congruence is attractive, it is not yet sufficiently developed to be operationalised with empirical methods due to assumptions related to typical realization, cf. Halliday's (2006:114) clarification, "a meaning that has usually been realised in one way is now being realised in another". The newness of *Brexit* prevents us from determining its 'usual' realization. However, it is clear from the discussion above that *Brexit* denotes an event and as such is an abstract nominal. While it might be difficult to make any strong claims about its (in)congruent realization, we have been able to see clearly that its construal is that of a hypothetical event with negative semantic prosody. With the Brexit clock ticking in the UK at the time of writing, we will have to wait to see what future developments of this lexeme lie in store.

## References

- Bains, Inderdeep. 2016. Brexit is named the word of the year by the Collins Dictionary after dominating politics in 2016, *The Daily Mail*, London, 3 November 2016. <https://www.dailymail.co.uk/news/article-3899852/Brexit-named-word-year-Collins-Dictionary-dominating-politics-2016.html> [accessed: 21 January 2019]
- Banks, David. 2008. *The Development of Scientific Writing, Linguistic Features and Historical Context*. London, Equinox.
- Barber, Charles. 1997. *Early Modern English*. New ed. Edinburgh: Edinburgh University Press.
- Bauer, Laurie. 1983. *English Word-Formation*. Cambridge University Press.
- Benczes, Réka. 2006. *Creative compounding in English: The semantics of metaphorical and metonymical noun-noun combinations*. Amsterdam and Philadelphia: John Benjamins.
- Clark, Herbert. 1974. Semantics and comprehension. In T. A. Sebeok (Ed.), *Current trends in linguistics: Linguistics and adjacent arts and sciences* (Vol. 12). The Hague: Mouton, pp. 1291-1428.
- Ebeling, Signe. 2018. Exploring Cross-Linguistic Congruence: The Case of Two Stance Frames in English and Norwegian. *Bergen Language and Linguistics Studies* 9 (1). <https://doi.org/10.15845/bells.v9i1.1419>.
- Fontaine, L. 2017a. Fluid boundaries and the categorization of nominal expressions. In S. Neumann, R. Wegener, J. Fest, P. Niemietz and N. Hützen (eds). *Challenging Boundaries in Linguistics: Systemic Functional Perspectives*. Aachen British and American Studies Series. Frankfurt am Main: Peter Lang Verlag, 19-38.

- Fontaine, Lise. 2017b. The early semantics of the neologism BREXIT: a lexico-grammatical approach. *Functional Linguistics*, 4:6. DOI: 10.1186/s40554-017-0040-x
- Goatly, Andrew. 1997. *The Language of Metaphors: Literal Metaphorical*. New York: Taylor & Francis.
- Halliday, Michael. 1966/2003. Grammar, society, and the noun. Reprinted in Michael Halliday and Jonathan Webster. *On Language and Linguistics. Collected Works of MAK Halliday*, Vol. 3. London: Continuum, 50-73.
- Halliday, Michael. 1994/2002. On grammar and grammatics. In J. Webster (ed.) *The Collected Works of MAK Halliday. Volume 1. On Grammar*. London: Bloomsbury Academic, pp 384-417.
- Halliday, Michael. 2004a. On the Grammar of Scientific English. In Jonathan Webster (ed.) *The Language of Science* (Volume 5 in the Collected Works of M. A. K. Halliday) London: Continuum, 181-198.
- Halliday, Michael. 2004b. Language and Knowledge: the 'Unpacking' of Text. In Jonathan Webster (ed.) *The Language of Science* (Volume 5 in the Collected Works of M. A. K. Halliday) London: Continuum, 24-48.
- Halliday, Michael. 2006/2013. Written Language, Standard Language, Global Language. In Jonathan J. Webster (ed.) *Halliday in the 21st century* (Collected Works of M.A.K. Halliday 11). London: Bloomsbury, pp 100-119.
- Halliday, Michael. 2007. On the concept of educational linguistics. In Jonathan Webster (ed.) *Language and Education* (Volume 9 in the Collected Works of M.A.K. Halliday). London: Continuum, 354-367.
- Halliday, Michael. 2010/2013. Language evolving: some systemic functional reflections on the history of meaning. In Jonathan J. Webster (ed.) *Halliday in the 21st century* (Collected Works of M.A.K. Halliday 11). London: Bloomsbury, 237-253.
- Halliday, Michael & Matthiessen, Christian. 2014. *An Introduction to Functional Grammar*. 4<sup>th</sup> edition. London: Arnold.
- Hansen-Schirra, Silvia, Neumann, Stella & Steiner, Erich. 2012. *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*. Berlin: Walter de Gruyter.
- Hasan, Ruqaiya. 2010. The meaning of 'not' is not in 'not'. In A. Mahboob and N. Knight (eds) *Applicable Linguistics*. London and New York: Continuum, 267-306.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubiček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel. 2014. The Sketch Engine: ten years on. *Lexicography*, 1, 1, pp 7-36.  
<http://www.sketchengine.co.uk>. Accessed 7 December 2016.

- Langacker, Ronald. 1987a. Nouns and Verbs, *Language*, 63 (1): 53–94
- Langacker, Ronald. 1987b. *Foundations of Cognitive Grammar*, Volume 1, Theoretical Prerequisites. Stanford: Stanford University Press.
- Matthiessen, Christian M.I.M., Teruya, Kazuhiro and Lam, Marvin. 2010. *Key terms in systemic functional linguistics*. London: Continuum
- xford English Dictionary "exit, v. and n." ●ED ●online. ●xford University Press, September 2015. Web. 2 November 2015.
- xford English Dictionary "fire, n. and int." ●ED ●online. ●xford University Press, December 2015. Web. 21 February 2016.
- xford English Dictionary "refusal, n." ●ED ●online. ●xford University Press, December 2015. Web. 21 February 2016.
- Plag, Ingo. 2003. *Word-formation in English*. Cambridge: Cambridge University Press
- Steiner, Erich. 2004. Ideational grammatical metaphor: Exploring some implications for the overall model. *Languages in Contrast*, 4, 1, 137–164.
- Taverniers, Miriam. 2011. The syntax–semantics interface in systemic functional grammar: Halliday's interpretation of the Hjelmslevian model of stratification. *Journal of Pragmatics*, 43, 1100–1126.
- Wegener, Rebekah. (2011). Parameters of context: From theory to model and application. Unpublished PhD thesis. Macquarie University, Sydney.





# CHAPTER TEN

## LANGUE NATURELLE, LANGAGE ICONIQUE ET SYMBOLIQUE

### NATURAL, ICONIC AND SYMBOLIC LANGUAGE

ANCA CHRISTINE PASCU

#### Abstract

The relationship between language and thought has been studied since antiquity. With the emergence of cognitive sciences at the end of the 20th century, a new element, knowledge, was added to the language-thought dyad and analysed along three dimensions: knowledge construction, knowledge transmission and knowledge acquisition.

In this article, we analyse the relationship between iconic and symbolic languages through knowledge in two ways. The first axis is that of knowledge construction, as a subsystem of the knowledge system. The second follows two major disciplines that have developed over the centuries: linguistics, whose focus is language and natural languages and logic, whose focus is thought and reasoning. Paralleling the distinction between linguistics and logic as disciplines, we can distinguish iconic language from symbolic language. Our aim is to establish some characteristics of these two types of language and their impact on the construction of knowledge.

#### 1. Introduction

Les rapports entre le langage et la pensée étudiés par la logique et par la linguistique en tant que sciences de l'homme, la première portant sur la pensée de l'homme, la seconde sur le mode d'expression de cette pensée, font l'objet d'étude de plusieurs disciplines notamment par la philosophie

du langage, la psychologie cognitive, l'informatique théorique et appliquée ou les mathématiques appliquées. L'intérêt de l'étude du rapport « pensée-langage » développé au sein de chacune de ces sciences a été, sans doute, double : d'une part, pour le développement des nouvelles « connaissances » propres à chaque discipline et, d'autre part, pour la transmission de ces connaissances. La langue naturelle a été longtemps le véhicule de transport des connaissances.

La logique en tant que discipline scientifique se développe à partir de l'antiquité comme la science du raisonnement, de l'analyse de la pensée, mais aussi de l'analyse du langage en tant que langue naturelle ainsi que de la prise de décisions. Vingt siècles après Socrate, Platon et Aristote, est apparue la Logique de Port Royal (1662 [Arnauld et Nicole 1964]), puis deux siècles plus tard (Frege 1971a ; Russell 1989 ; Whitehead [Vernant 2009]) la création du *langage symbolique* de la logique mathématique moderne. Ce langage symbolique, qui pourrait fonctionner comme un modèle au sens de la théorie des modèles, a pourtant effacé un nombre de traits cognitifs conceptuels en privant la logique moderne (le calcul du premier ordre [Kleene 1971]) et ses extensions d'une large étendue en ce qui concerne ses applications. C'est à partir de cette faiblesse qu'une partie des logiques dites déviantes se sont développées. Les autres, comme la Logique de la Détermination des Objets (LDO) (Desclés et Pasco 2011) ont renoué avec certains concepts anciens en les traduisant dans le langage symbolique de la logique moderne. On peut dire que, maintenant, l'objet de la logique n'est plus uniquement l'analyse du raisonnement comme il l'était traditionnellement, mais aussi les objets du raisonnement, leur construction par l'esprit humain, les systèmes de prise de décisions et d'autres aspects cognitifs qui n'étaient pas pris en compte par la logique des XIX<sup>e</sup> et XX<sup>e</sup> siècles (Frege 1971a ; Russell 1937).

En ce qui concerne la linguistique, il y a une particularité épistémologique de cette discipline – son objet d'étude est représenté par elle-même. Selon David Banks (2011 : 265) :

La linguistique est une discipline bien particulière, dans ce sens où elle est la seule discipline où objet d'étude et son moyen d'expression coïncident. La linguistique étudie la langue, mais l'étude de la langue ne peut s'exprimer qu'à travers une langue, comme le dit M. Halliday, le fondateur de la Linguistique Systémique Fonctionnelle.

Dans le même ouvrage, D. Banks va plus loin en citant M. Halliday<sup>1</sup> :

---

<sup>1</sup> Les citations ont été traduites par Banks (2011)

- All systematic knowledge takes the form of 'language about' some phenomenon; but whereas the natural sciences are language about nature, and the social sciences are language about society, linguistics is language about language « language turned back on itself », in Firth's often quoted formulation. (Halliday 1996 (2002) : 384)
- Language is a part of the social system. (Halliday 1978 : 39)
- [A] text is a sociological event, a semiotic encounter through which the meanings that constitute the social system are exchanged. (Halliday 1978, 139.)
- A text, then, is both an object in its own right ... and an instance an instance of social meaning in a particular context of situation. (Halliday & Hasan 1989, 11).

M. Halliday et les linguistes du courant de linguistique systémique fonctionnelle partent de trois caractéristiques du langage pour définir le modèle de la linguistique systémique fonctionnelle :

- L'objet de l'étude du langage est le langage.
- Le langage est un système au sens de la théorie des systèmes.
- Le langage est un système social créé dans le contexte et par des humains.

La linguistique systémique fonctionnelle est un modèle linguistique qui interconnecte dans un même système trois dimensions du langage : la capacité du langage de créer des représentations des réalités du monde (la métafonction idéationnelle), la capacité du langage de permettre aux humains de créer de relations sociales (la métafonction interpersonnelle) et la capacité du langage de créer et d'élaborer le discours (la métafonction textuelle). Dans cette caractérisation du langage apparaissent ainsi des

- 
- Toute connaissance systématisée prend la forme d'un « langage au sujet » d'un phénomène quelconque ; mais si les sciences naturelles constituent un langage au sujet de la nature, et les sciences sociales constituent un langage au sujet de la société, la linguistique est un langage au sujet du langage – le langage replié sur lui-même. Comme le dit la citation bien connue de J.R. Firth.
  - La langue fait partie du système social.
  - [...] un texte constitue un événement sociologique, une rencontre sémiologique à travers laquelle les significations qui constituent le système social sont échangées.
  - Un texte est un objet en soi ... et une occurrence – une occurrence de signification sociale dans une situation particulière.

éléments formels qui restent très proches de la langue naturelle par sa fonction de système social.

L'analyse du discours est sans doute une dimension très importante de la linguistique, la linguistique informatique apparaissant dans un dernier temps. Elle construit des modèles qui sont à la base d'une analyse automatique (par la machine) du langage et, donc, de la langue. C'est aussi pour l'analyse du rapport entre le raisonnement et le langage que certains modèles sont construits. Dans ce but, on considère d'autres dimensions du langage qui sont exprimées par les représentations du langage iconique et du langage symbolique. Ces deux types du langage sont plus proches de la modélisation pour une implémentation informatique que de la langue naturelle.

Dans cet article, nous proposons une analyse du langage iconique et symbolique situés dans le contexte de la construction, l'acquisition et la transmission des connaissances par le langage et dans le contexte de l'évolution des notions d'*objet* et de *concept*, notions premières dans toute analyse logique ou linguistique du langage.

Si l'on veut analyser le langage iconique par rapport au langage symbolique et mieux cerner une définition pour ces deux types de langage, il faut partir de deux notions : *sens* et *représentation*. Ces deux notions ont fait l'objet d'étude de la philosophie du langage et, plus tard, des sciences du langage et des sciences cognitives. Les acquis actuels en intelligence artificielle et en informatique appliquée montrent que la frontière stricte entre l'iconique et le symbolique ne peut être tracée ; cependant, les deux sont dépendants du domaine de la connaissance. A partir des hypothèses sur ce qui est la *représentation d'une connaissance*, des questions se posent sur ces deux langages :

- Peut-on voir l'iconique comme une forme de représentation des connaissances ?
- Peut-on voir le symbolique comme une autre forme de représentation de connaissances ?
- Quel serait le rapport entre ces deux formes ?

Notre hypothèse est que ces deux types de langages représentent deux formes de connaissances et que le symbolique peut être dérivé de l'iconique par un processus de *compilation* par analogie à la compilation des langages informatiques.

La structure de cet article est la suivante : après une discussion générale présentée en introduction, nous analysons, en deuxième partie, le rapport entre le langage, la pensée et la construction des connaissances. La troisième

partie traite du langage conçu comme un système de construction des connaissances ; elle présente un bref historique de l'évolution des notions de concept, objet, sens, dénotation, notions primitives dans l'analyse du langage en général et dans le langage iconique et symbolique en particulier. La quatrième partie analyse le rapport entre la langue naturelle, le langage iconique et le langage symbolique et propose la caractérisation de ces deux types de langage.

## 2. Langue naturelle, langage, pensée et construction des connaissances

La frontière entre langue naturelle et langage n'est pas nette. La caractéristique principale qui les distingue est la suivante : si la langue naturelle est représentée par l'ensemble des langues produites par l'humanité, le langage est construit, résultat d'un processus de raisonnement à partir d'une langue naturelle ayant comme but la construction et l'organisation des connaissances.

La pensée couvre un champ plus large ayant comme composantes au moins le raisonnement et la décision. Comment peut-on définir la construction des connaissances ? Tout d'abord, nous devons adopter une définition de la *connaissance*. Parmi les nombreuses définitions, C. Godin (2004) propose la suivante :

1. Faculté mentale produisant une assimilation par l'esprit d'un contenu objectif préalablement traduit en signes et en idées.
2. Résultat de cette opération. La connaissance est une possession symbolique des choses. Elle comprend une infinité de degrés. La connaissance rationnelle, méthodique universelle a parfois été opposée au savoir empirique, chaotique, objectif.

Même si cette définition n'est pas donnée dans le cadre d'un système formel, d'une manière formalisée, elle capte les traits cognitifs de la connaissance et, par son point 1, exprime la *construction des connaissances*. Cette définition met en évidence deux aspects cognitifs : la construction vue comme une opération mentale et le résultat vu plutôt comme un objet ayant différentes formes et types. Cet objet entre dans d'autres opérations cognitives comme l'*acquisition* et la *transmission*. On peut considérer un système de connaissances étant composé de trois sous-systèmes : *sous-système de construction*, *d'acquisition* et *de transmission*. Les éléments structurels d'un système de connaissances sont *les objets* et *les concepts*.

Notre analyse se situe dans le cadre de ces trois sous-systèmes mis en rapport avec les deux grandes sources de connaissances : *la langue* et *la pensée*. Elle veut montrer que la construction des connaissances et, donc, la transmission et l'acquisition des connaissances se font à partir de la langue naturelle, en passant par le langage, soit le langage iconique, soit le langage symbolique, soit les deux. Si la connaissance est traitée d'une manière automatique (par la machine), elle doit nécessairement être l'objet d'un système (langage) formel.

Un *langage formel* est un ensemble formé de deux éléments : un sous-ensemble d'*éléments primitifs* et un ensemble de *règles de composition*. Les primitives sont en lien étroit avec un sens, donné par les *concepts* sous-jacents. Ces derniers, reliés en réseau (*réseau des concepts*), peuvent être considérés comme un *support du sens*. Leur lien avec le sens est de *nature structurelle*. Le système de règles est lié aussi au sens, mais ce lien est plutôt de *nature fonctionnelle*.

Une instanciation de cette définition du langage peut être retrouvée dans les langues naturelles. Par analogie avec le fonctionnement des langues naturelles où les primitives sont les *mots* (ou une autre unité primitive comme les *caractères* dans certaines langues asiatiques<sup>2</sup>, le système de règles est le *système syntaxique*. Le lien fonctionnel entre les deux s'exprime, dans une manière formalisée, par une extension possible du *principe de la compositionnalité* de la logique classique. Autrement dit, si l'on traite les connaissances comme un système, il faut traiter le langage comme un système ; la linguistique en est le modèle, en tant que science.

Si la pensée représente le raisonnement, alors le modèle du raisonnement est la logique.

### 3. Langage, système de construction de connaissances

Dans cette partie, nous présentons un court historique de l'évolution des notions de *concept*, *objet*, *sens* et *dénotation* grâce à deux points clés dans l'histoire de la logique et de la philosophie représentés par Frege et Russell. Frege et Russell sont aussi les précurseurs des mathématiques modernes et du langage symbolique moderne en mathématiques. Les notions susmentionnées sont des notions primitives dans l'analyse du langage, y compris dans l'analyse de ses types comme le langage iconique et le langage symbolique.

Une analyse du langage et des langues en tant que systèmes de construction de connaissances et, par conséquent, le statut du langage

---

<sup>2</sup> Eduardo Fazzoli, *Caractères chinois (en 214 clés)*, Ed. Flammarion, Paris, 1987



iconique et le langage symbolique nécessitent, d'abord, de cerner le statut notionnel de quelques notions primitives comme : *sens*, *dénotation*, *représentation* et *image*.

### 3.1 Sens, dénotation, représentation et image

Une analyse du rapport entre le sens et la dénotation passe nécessairement par la représentation. La première analyse systémique<sup>3</sup> des notions de *sens* et de *dénotation* est proposée par Frege dans son système conceptuel destiné à être un fondement logique pour les mathématiques (Frege, 1971a, 1971b).

En ce qui concerne la notion de *représentation*, Peirce (1995 [1898]) utilise un symbolisme basé sur des diagrammes pour décrire la logique du premier ordre<sup>4</sup>. Il a construit une méthode diagrammatique permettant de faire des déductions dans le calcul propositionnel et des prédicats. Le caractère déductif d'une structure est proche d'une intuition géométrique pour Peirce. Cette analogie permet dans une certaine mesure de donner une preuve de la valeur de vérité d'une proposition.

En ce qui concerne l'image, il n'existe pas une définition unique. La définition suivante, proposée par Wikipedia est assez évocatrice quant à la complexité de l'interprétation :

Une image est une représentation visuelle voire mentale de quelque chose. Elle peut être naturelle ou artificielle, visuelle ou non, tangible ou conceptuelle, elle peut entretenir un rapport de ressemblance directe avec son modèle ou au contraire y être liée par un rapport plus symbolique. (<http://fr.wikipedia.org/wiki/Image>)

Les travaux plus récents abordent l'image du point de vue de chaque domaine ou sous-domaine qui est concerné par l'image. Gottlob Frege (1848–1925) est le premier à rompre avec la tradition qui a traversé le Moyen Age en analyse du langage. La logique scolastique du Moyen Age continue dans la tradition aristotélicienne à considérer les propositions comme étant catégoriques et hypothétiques. Les propositions catégoriques sont particulières et universelles. Les propositions hypothétiques sont

<sup>3</sup> En analyse systémique, le mot « systémique » est utilisé au sens d'une analyse réalisée dans le cadre d'un système, en occurrence le système de Frege.

<sup>4</sup> La logique de premier ordre ou la logique classique ou encore le calcul des prédicats du premier ordre sont les noms sous lesquels est connue la logique classique.

nécessaires, contingentes, possible, impossible, vraies, fausses. La valeur de vérité est indifférente à la modalité (Busquets 2006).

La dichotomie catégorèmes-syncatégorèmes dans l'analyse des composantes d'une proposition prend une dimension plutôt linguistique. Chez William of Sherwood (1205-1270) et Pierre D'Espagne (1220-1277) apparaît la différence entre catégorèmes (nom et verbes) et syncatégorèmes (prépositions, conjonctions, etc.)

Pour Guillaume d'Occam (1285-1347), les syncatégorèmes n'ont pas de signification définie mais ils peuvent modifier la signification d'un terme (Biard 1989).

Plus tard, Husserl (1859-1938), dans son système logique d'analyse du langage, considère que les catégorèmes ont une signification par elles-mêmes, tandis que les syncatégorèmes n'ont une signification qu'en association avec d'autres termes (Husserl 1969).

Le statut des catégorèmes, des syncatégorèmes et leur emploi dans un modèle sémantique du langage a produit un vrai débat parmi les philosophes du langage comme Dummett, Quine, Sommers entre 1960 et 1980.

Dans toutes ces approches du sens, deux idées émergent. La définition de l'opposition catégorème-syncatégorème est donnée par des moyens qui tiennent de la linguistique plutôt que de la logique ; l'indépendance par rapport à la dépendance en ce qui concerne la signification n'est pas la même pour les uns et pour les autres.

Frege s'appuie sur la notion mathématique de *fonction* pour caractériser les composantes d'une proposition. Il oppose la fonction, qui est pour lui une entité non saturée, à l'objet qui est une entité saturée. Pour la première fois, il donne aussi la définition mathématique d'un *concept* comme étant une fonction qui s'applique à un objet en rendant une valeur de vérité comme résultat.

Pour fonder son système logique, Frege (1971a) utilise des termes spécialisés, regroupés par paires :

- Nom : sens et dénotation ;
- Fonction : fonction et objet (non-saturé opposé à saturé) ;
- Nom et étiquette ;
- Fonction et parcours de valeurs ;
- Nom (étiquette) des valeurs de vérité et assertion.

Frege commence son analyse par la description des noms. Un nom complet nomme un objet, par exemple, Napoléon Bonaparte ou Vénus. Un nom incomplet nomme une fonction, par exemple, le frère aîné de [...]. Un nom simple est de type Napoléon Bonaparte, Vénus, Nessus (le centaure),

9 (le nombre). Un nom complexe : le frère aîné de Napoléon Bonaparte, la chemise de Nessus.

La dénotation d'un nom propre est l'objet même que nous désignons par ce nom ; la représentation que nous y joignons est entièrement subjective ; entre les deux se trouve le sens, qui n'est pas subjectif comme l'est la représentation, mais qui n'est pas non plus l'objet lui-même.

Tous ces commentaires se trouvent dans l'introduction rédigée par M. Furth à l'œuvre de Frege (1967 : 5).

Les axiomes de Frege sont les suivants :

1. Un nom *exprime* un sens et *dénote* une dénotation ;
2. Une certaine relation s'établit entre sens et dénotation ;
3. Le sens d'un nom complexe est une fonction des sens de ses composantes ;
4. La dénotation d'un nom complexe est une fonction des dénotations de ses composantes ;
5. Un nom peut exprimer un sens, mais il peut manquer de dénotation ; pas l'inverse ; par exemple, le nom Ulysse exprime un sens, mais il n'a pas de dénotation ;
6. Deux expressions complexes ayant la même dénotation peuvent exprimer des sens différents (exemple : pour les expressions mathématiques  $2^2$  et  $2+2$ , la dénotation est la même, le nombre 4, mais le sens est différent ; la première expression exprime une puissance, la seconde une addition) ;
7. La dénotation d'un nom complet est un *objet*.

Frege (1971a) associe au système sens – dénotation la *représentation*. La représentation associée à un signe doit être distinguée de la dénotation et du sens de ce signe. Si un signe dénote un objet perceptible au moyen des sens, ma représentation est un tableau intérieur, formé du souvenir des impressions sensibles et des actions externes ou internes auxquelles je me suis livré. Chez le même individu, la même représentation n'est pas toujours liée au même sens. Il est bien naturel que les représentations associées au même sens diffèrent grandement entre elles.

Un peintre, un cavalier et un naturaliste lieront sans doute des représentations bien différentes au nom Bucéphale. C'est par là qu'une représentation se distingue essentiellement du sens d'un signe. Celui-ci peut être la propriété commune de plusieurs individus, il n'est donc pas partie ou mode de l'âme individuelle. Car on ne pourra pas nier que l'humanité possède un trésor commun de pensées qui se transmet d'une génération à une autre.

C'est pourquoi il ne convient pas de désigner par le mot « représentation » ce qui en diffère si profondément. Cette idée rejoint ce que l'on appelle aujourd'hui la *représentation sociale*.

Un schéma du système frégeén du rapport entre sens (concept) et dénotation (objet) (Pascu 2001) est présenté dans la figure 1.

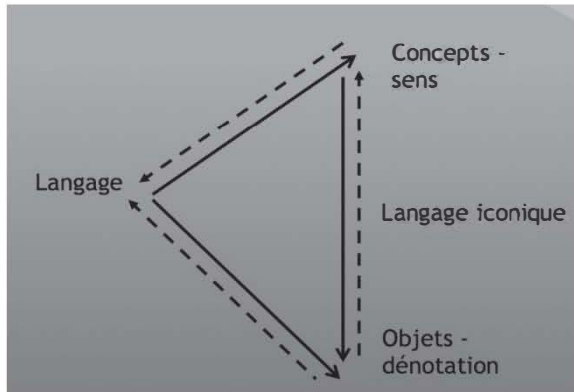


Figure 1. Sens et dénotation chez Frege

Le sens pour Frege est une preuve d'existence, existence prise dans le sens de « preuve réelle ». L'image dans la lunette existe, tandis qu'Ulysse n'existe pas. On peut admettre comme preuves d'existence d'autres types de preuves, par exemple :

- pour l'image de la lune dans la lunette – mesure,
- pour Ulysse – le mythe,
- pour le nombre 4 – la construction arithmétique des nombres naturels.

Dans ce cas, l'axiome 5 de Frege ci-dessus devient discutable. En fait, l'axiome 5 de Frege affirme qu'entre sens et dénotation existe une correspondance qui n'est pas biunivoque. Si l'on admet d'autres preuves de l'existence que l'existence réelle, alors cette correspondance devient biunivoque et le sens et la dénotation sont équivalents. Ce n'était pas le cas pour Frege.

C'est une des raisons pour lesquelles, d'autres systèmes logiques ont étendu la problématique du sens et de la dénotation en l'exprimant par la « dualité » *concept – objet*. La logique qui est fondée sur ces deux notions est la Logique de la Détermination des Objets (LDO) (Desclés et Pascu 2011). Elle s'inscrit dans la tradition de pensée de la Logique de Port-Royal

(Arnauld et Nicole 1992 ; Auroux, 1993). La logique de Port-Royal postule les traits cognitifs d'un *objet* et d'un *concept*, la notion d'*extension* et d'*intension*, la notion de détermination sans, évidemment, les formaliser. La LD● les formalise, les intègre dans un formalisme logico-mathématique et, en même temps, donne un statut dans le cadre de la logique à d'autres caractéristiques cognitives comme la *typicalité / atypicalité*.

La Logique de la Détermination des ●bjets repose sur deux notions primitives : concept et objet. Ces deux notions donnent un statut de pré-logique à la Logique de la Détermination des ●bjets. Cette logique contient une « description » des objets sur lesquels on opère des inférences. Ces primitives sont les objets et les concepts. Les propositions et les prédicats s'obtiennent en appliquant des concepts aux objets.

La définition d'un concept part de l'idée que les concepts sont liés dans des réseaux de concepts par la relation d'*héritage* (héritage des propriétés – une propriété complexe est définie à l'aide des propriétés plus primitives qu'elle).

En LD●, la définition d'un objet est enrichie par rapport à Frege. Pour Frege, l'objet était défini comme une *entité saturée*, par opposition à la fonction qui est une *entité non-saturée*. En LD●, les objets sont de deux types : *objets totalement déterminés* qui coïncident avec les objets frégeïens et *objets plus ou moins déterminés*. Les objets totalement déterminés associés à un concept f forment l'extension de f (Ext f). L'extension est prise au sens de la logique classique. Les objets plus ou moins déterminés s'obtiennent à partir d'un objet de pensée associé à chaque concept appelé *l'objet typique totalement indéterminé* et dénoté par  $\tau f$  par des déterminations successives. L'opération de *détermination* est une autre primitive de la LD●. L'ensemble de tous les objets plus ou moins déterminés associés à un concept constitue l'étendue du concept f (Etendue f).

Le concept est un quintuple formé de la propriété même, f, d'un paquet de propriétés sous-jacentes à f, l'intension de f, Int f, d'un sous-ensemble de l'intension représenté par toutes les propriétés strictement nécessaires à la propriété f (les propriétés sans lesquelles f ne peut pas être f), (Ess f) et les deux ensembles d'objets l'étendue (Etendue f) et l'extension, Ext f).

$$C = (f, \text{Int } f, \text{Ess } f, \text{Etendue } f, \text{Ext } f)$$

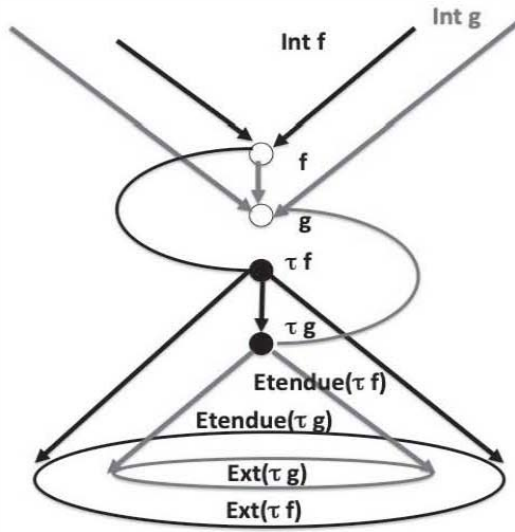


Figure 2. Le système conceptuel de la LDO●

Le système conceptuel de la LDO est représenté dans la figure 2.

Russell (1989, 86) donne une autre interprétation à la dénotation, plus liée à la proposition et à sa valeur de vérité :

La notion de dénotation, comme la plupart des notions logiques, s'est trouvée, jusqu'ici, obscurcie par l'intervention de considérations psychologiques hors de propos. En un certain sens, nous dénotons quand nous indiquons ou nous décrivons, ou quand nous utilisons les mots comme des symboles de concepts ; tel n'est pas, cependant, le sens que je souhaite examiner. Le fait que la description soit possible – que nous soyons capables, au moyen de concepts, de désigner une chose qui n'est pas un concept – est dû à une relation logique entre certains concepts et certains termes, en vertu de laquelle ces concepts dénotent intrinsèquement et logiquement ces termes.

Pour Russell, une expression dénotante est (1989, 203) :

[...] Une expression semblable à n'importe laquelle des expressions suivantes : un homme, quelques hommes, n'importe quel homme, chaque homme, tous les hommes, l'actuel roi de France, le centre de la masse du



système solaire au premier instant du XX<sup>ème</sup> siècle, la révolution de la Terre autour du Soleil, la révolution du Soleil autour de la Terre<sup>5</sup>.

Il distingue les cas suivants :

1. expression dénotante qui ne dénote rien : l'actuel roi de France ;
2. expression dénotante qui dénote un objet déterminé : l'actuel roi d'Angleterre ;
3. expression dénotante qui dénote d'une manière ambiguë : un homme

La dénotation est liée à la quantification [Russell, 1902 – 1918, 205]:

C(tout) : « C(x) est toujours vrai »

C(rien) : « “C(x) est faux” est toujours vrai »

C(quelque chose) : « il est faux que “C(x) est faux” est toujours vrai »

Les notions primitives de la logique de Russell, notions qui constituent le fondement logique des mathématiques modernes, sont les suivantes :

- la notion de variable,
- la notion logique de proposition,
- la notion de valeur de vérité,
- l'identité,
- la négation,
- la notion de classe – fonction, extension, classe.

Le principe de l'extensionnalité – deux classes sont identiques si et seulement si elles ont les mêmes membres, deux fonctions sont équivalentes si elles déterminent la même classe :

$$[(\forall x) (\varphi x \equiv \psi x)] \Leftrightarrow [\hat{c}(\varphi x) \equiv \hat{c}(\psi x)]$$

L'analogie syntaxique : « le roi de France », « le roi d'Angleterre »

Le symbolisme du langage logique efface un bon nombre de traits conceptuels – les traits psychologiques (émotions, attitudes...). Les axiomes de la logique classique ont été revisités par les logiques modernes dites déviantes en ce qui concerne la remise en cause de : l'identité, la négation, la dualité extension-intension, la notion d'opérateur par rapport à la notion de fonction, les notions d'objet et de concept.

Frege et Russell sont considérés comme les fondateurs des mathématiques modernes. Frege est le premier qui donne une définition formelle au « concept » et à l'« objet ». Russell fait une analyse du « concept » et de sa « dénotation ». Par le biais de cette analyse, il crée un langage

---

<sup>5</sup> Les citations ci-dessus appartiennent à la traduction française de Russell.

symbolique qui, avec quelques modifications, reste aujourd'hui le langage symbolique des mathématiques.

#### 4. Langue naturelle, langage iconique, langage symbolique

Toute la connaissance de l'humanité a été construite, acquise et transmise de la nuit des temps par les langues naturelles. Plus tard, la structuration des connaissances en systèmes a imposé la construction d'un langage en adéquation avec les objets d'étude. Ce langage est, dans un premier temps, de type iconique (dessins, schèmes,...) et évolue vers un symbolisme de haut niveau comme c'est le cas en mathématiques modernes ou physique théorique.

Ensuite, chaque science a construit son langage dépendant dans une certaine manière de ses objets d'étude. C'est le langage symbolique.

Pour pouvoir cerner une définition de ces deux types de langage, nous fournissons quelques exemples ; ils sont tirés des mathématiques et de la linguistique formelle. On peut considérer les mathématiques comme étant la première science à développer son langage symbolique ; la linguistique formelle, repose sur la modélisation mathématique.

En mathématiques, arithmétique et algèbre, les objets primitifs sont les *nombre*s. Le langage porte, donc, sur les nombres. Prenons quelques exemples de problèmes mathématiques traités en utilisant le langage iconique et, en parallèle, le langage symbolique. Tous ces exemples tiennent de la construction des connaissances.

**En arithmétique** : Quel est le nombre auquel en rajoutant 3, on en obtient 5.

« Le nombre auquel en rajoutant 3 on obtient 5 est 2 ».

Une solution possible en arithmétique, en n'utilisant que le langage iconique, est présentée dans la figure 3.

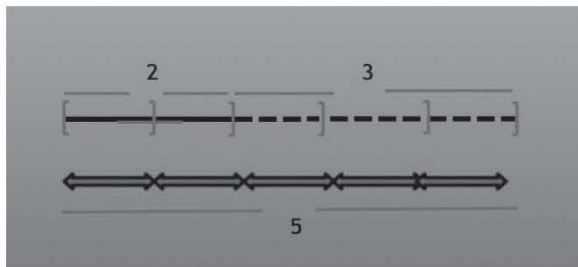


Figure 3. Résolution arithmétique de l'équation :  $x + 3 = 5$

**En algèbre :** On cherche la solution de l'équation  $x+3 = 5$ .

Cette solution est  $x = 2$

Nous remarquons que même si on travaille avec des objets définis à l'intérieur des mathématiques, notamment les nombres, en arithmétique on utilise la langue naturelle tandis qu'en algèbre on utilise un langage symbolique qui impose la notion de *variable*. Dans ce cas, la variable est dénotée par  $x$ .

Pour la résolution de l'équation de premier degré, le langage iconique n'est plus utilisé sauf dans les manuels d'arithmétique de l'école primaire où la solution est trouvée intuitivement par l'image de la figure 3.

**En géométrie synthétique :** Les figures géométriques sont représentées dans un langage iconique (Figure 4).

**En géométrie analytique :** Les mêmes figures sont représentées par leurs équations, donc on utilise un langage symbolique.

Dans la figure 4, nous trouvons la représentation du cercle ayant le centre en origine et le rayon 1,5. Son équation donnée en coordonnées polaires est :  $x^2+y^2 - 1,5^2 = 0$ .

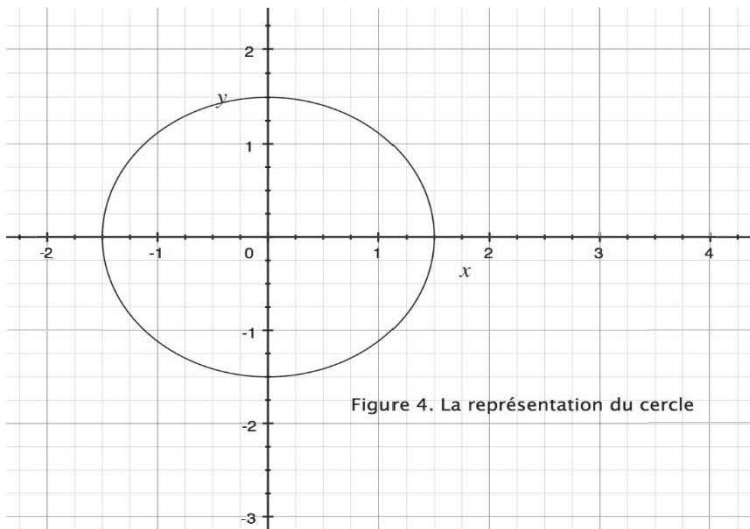


Figure 4. La représentation du cercle

Figure 4. La représentation du cercle

Si l'on reste dans le domaine des mathématiques et, notamment en géométrie algébrique, il existe un concept qui correspond à une surface nommée *tore*. La définition en langue naturelle du *tore* est donné par Wikipédia, <http://en.wikipedia.org> :

In mathematics, a toroid is a surface of revolution with a hole in the middle, like a doughnut<sup>6</sup>.

Dans la figure 5, on peut trouver la représentation du tore en langage iconique (l'image) et en langage symbolique (l'équation). Nous pouvons remarquer qu'au niveau de la compréhension, la distance entre la langue naturelle et les deux langages iconique et symbolique est assez grande. En revanche, la distance entre le langage iconique et le langage symbolique reste assez étroite. Autrement dit, pour quelqu'un qui n'a pas une formation en géométrie algébrique, il est difficile de comprendre ce qu'est un tore à partir de son image, et encore plus difficile à partir de son équation.

$$\begin{cases} r_0 \\ \theta \\ z \end{cases} = \begin{cases} 3+\sin u+\cos u \\ 2t \\ \sin u+2\cos t \end{cases}, t=0, \dots, 2\pi, u=0, \dots, 2\pi$$

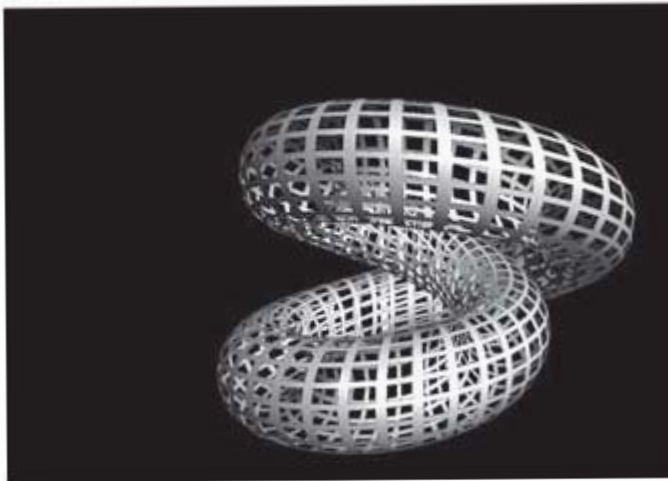


Figure 5. Le tore

En logique : La règle de déduction connue aussi sous le nom de *modus ponens* s'énonce en langue naturelle : Si une proposition  $p$  est vraie et on sait que cette proposition entraîne une autre proposition  $q$ , alors la proposition  $q$  est vraie

<sup>6</sup> En mathématique, un tore est une surface avec un trou au milieu, comme un beignet (donut).

En langage symbolique de la logique classique présentée sous la forme de la déduction naturelle, la règle de déduction s'écrit comme dans la figure 6.

Le langage symbolique de la logique actuelle est passé par une étape plus proche de l'iconicité.

La figure 7 présente les représentations de Peirce (1995 [1898]) pour les connecteurs et quantificateurs logiques : *non p* ; *p et q* ; *p ou q* ; *p implique q* ; *il existe x tel que F(x)* ; *quels que soit x F(x)*.

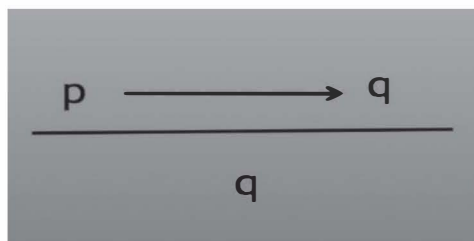


Figure 6. Modus ponens

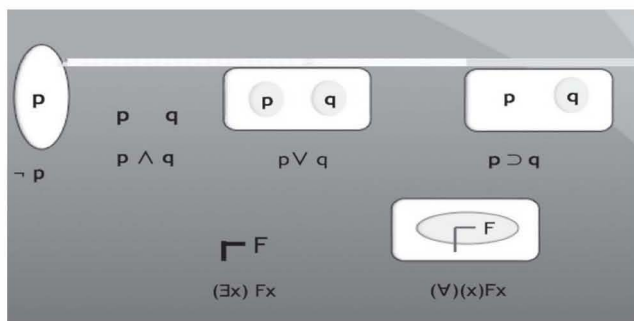


Figure 7. Les connecteurs logiques chez Peirce

La figure 8 contient la représentation de l'assertion d'une proposition chez Frege et la représentation symbolique de l'implication propositionnelle ( $A \supset B$ ).

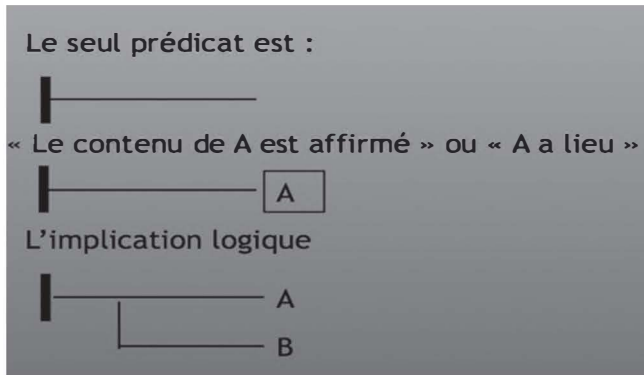


Figure 8. L'idéographie (Der Begriffsschrift, Frege, 1879)

Par ces exemples, on peut voir l'évolution du langage symbolique de la logique de premier ordre en passant par des étapes d'iconicité, mais qui ont été définies en strict lien avec le contenu des notions représentées.

Il faut rappeler ici que la logique de premier ordre (le calcul des propositions et des prédicats) ou la logique intuitionniste peut être présentée d'une manière symbolique sous forme de langage formel – appelé « voie profane » par Van Dalen, ou sous forme de déduction naturelle – appelé la « voie sacrée » par Van Dalen :

Logic appears in a “sacred” and in a “profane” form; the sacred form is dominant in proof theory, the profane form in model theory. (1997 : 1)

Les sciences formelles (les mathématiques, la physique théorique, la biologie théorique) se sont développées à partir de l'observation vers une généralisation exprimée par des lois, des théorèmes ou des corollaires. A cette fin, elles ont engendré des langages symboliques. Par rapport à la nature de leurs objets d'étude (les nombres ou des objets construits basés sur des nombres en mathématique, des objets construits basés sur des données expérimentales en physique...) ces langages contiennent des éléments d'iconicité (figures, schèmes...).

La situation n'est pas la même en ce qui concerne les sciences humaines. Pour ces dernières, le langage reste moins « formel »<sup>7</sup>. Pourtant, dans le cadre des sciences humaines, l'utilisation de l'informatique en traitement des données a imposé l'utilisation des modèles formels de nature

<sup>7</sup> Ici, on emploie le mot « formel » au sens de l'appartenance à un système formalisé, un système qui se soumet à des règles précises, des axiomes ...



mathématique, logico-mathématique ou statistique et, donc, l'utilisation d'un langage symbolique. Le langage symbolique doit décrire (modéliser) les traits cognitifs des objets et des concepts propres au système conceptuel de la science humaine dont il modélise le problème.

Le cheminement de résolution d'un problème est le suivant :

- ● On repère un « problème »<sup>8</sup> au sein d'une science humaine (science d'origine).
- ● On met en évidence ses caractéristiques en utilisant les concepts, et ainsi le langage, de la science d'origine.
- ● On transpose par analogie ces caractéristiques dans des concepts du modèle (mathématique, statistique, etc.). C'est ici qu'intervient le langage symbolique de la science auquel le modèle appartient (science de destination).
- ● On applique le traitement adéquat du modèle selon la science de destination (en utilisant, éventuellement l'ordinateur).
- ● On interprète le résultat fourni en termes du langage de la science de destination, en faisant une traduction des concepts dans le langage propre à la science d'origine.

Prenons le cas de la modélisation en théorie de la catégorisation : on voudrait regrouper en classes un nombre d'individus décrits selon un ensemble de propriétés. Si l'on applique un modèle mathématique de catégorisation, chaque individu devient un point dans un espace vectoriel de dimension égale au nombre de propriétés. L'individu qui a une certaine signification dans la discipline d'origine devient un élément d'un espace vectoriel en mathématique. Un modèle mathématique de catégorisation traite les données d'entrée et rend comme résultat un ensemble de classes qui regroupent les individus. L'interprétation de la pertinence de ces classes sera faite dans la discipline d'origine.

**En linguistique** : Une longue tradition de réflexion sur le langage qui remonte à l'Antiquité en passant par le Moyen Age s'est développée en suivant au moins trois perspectives : l'analyse de la structure du langage, l'analyse du rapport entre le langage et la pensée, l'analyse du langage comme objet social. Et pourtant, ce n'est qu'à partir du début du XX<sup>e</sup> siècle que l'on considère la linguistique comme étant une discipline scientifique. Une vraie démarche d'étude scientifique sur le langage a été commencée

---

<sup>8</sup> Ici, la notion de « problème » est comprise dans un sens élargi qui s'applique aussi bien aux sciences humaines qu'aux autres sciences. En linguistique, par exemple, un problème est la conceptualisation du temps et de l'aspect, en psychologie modéliser les émotions, en littérature, annoter un texte par rapport aux points de vue.

par Ferdinand de Saussure suivi par Emile Benveniste et, ensuite, par Roman Jakobson, André Martinet, Louis Hjelmslev. Ils sont considérés comme appartenant au structuralisme en linguistique.

Le structuralisme linguistique ne constitue pas une école unique se référant à une doctrine précise ; il s'agit bien plutôt d'un ensemble de courants, voire d'individus, qui ont élaboré des théories diverses, mais qui se fondent sur certains principes généraux communs. Le nom même de « structuralisme » indique que la langue est conçue comme une structure, c'est-à-dire comme un ensemble d'unités structurées par des réseaux de relations.

C'est en prenant le contre-pied de la démarche structuraliste que s'est constitué, vers la fin des années 1950, le courant le plus connu des *grammaires formelles*, la *grammaire générative transformationnelle* de Noam Chomsky.

Pour la linguistique, le modèle formel le plus saillant, autour duquel a été créé un nombre de modèles dérivés est celui de la *grammaire formelle*. Une *grammaire formelle* est le modèle formel obtenu par analogie de la grammaire d'une langue.

Plusieurs types de grammaires ont été développés en tant que modèles d'analyse du langage en mettant en avant dans le modèle certaines caractéristiques de la structure du langage. Au moins trois grandes catégories s'en détachent.

- La *grammaire générative* (Chomsky 1965). Le modèle thématise l'agencement et l'ordre des mots dans la phrase. Il est défini dans une perspective purement syntaxique. Ce modèle s'est avéré un bon outil pour l'analyse syntaxique des langages de programmation. Les langages formels associés à différentes classes de grammaires génératives ont trouvé des applications en biologie. Certains types de grammaires fondés sur le modèle générativiste ont été étendus en englobant d'autres paramètres linguistiques que l'ordre des mots. Ces grammaires ont été appliquées aux problèmes d'ingénierie linguistique comme « génération du langage » (*language generation*), « annotation sémantique du texte » (*text semantic annotation*), recherche de l'information (*information retrieval*) « résumé automatique » (*text summarising*) ou « traduction automatique » (Godin Christian, *Dictionnaire de philosophie, Paris, Fayard, 2004, 1534 p.*)

- Les différents types de grammaires catégorielles (Adjukiewicz, Bar Hillel, Lambek<sup>9</sup>, Steedman, Desclés, Biskri (Biskri et Desclés 1995)<sup>10</sup>. Le modèle est fondé sur l'ordre des mots et sur une catégorisation des unités syntaxiques sous forme de types syntaxiques. Dans les grammaires catégorielles, on retrouve l'ancienne idée de catégoriser les unités linguistiques selon la dichotomie catégorèmes–syncatégorèmes. En général, les catégorèmes correspondent aux types de base, les syncatégorèmes aux types composés. Les différentes grammaires catégorielles se distinguent par la définition des types : types syntaxiques ou types sémantiques. Le fonctionnement de cette grammaire est représenté par le calcul sur les types en suivant les *règles de la grammaire*<sup>11</sup>.

- La grammaire applicative universelle (Shaumyan 1987). La grammaire applicative et cognitive (Desclés, 1990). Ces grammaires sont des modèles conçus comme de systèmes composés de sous-systèmes dont la primitive de base est l'application d'un *opérateur* à un *opérande*.

Une *grammaire générative* est un quadruple  $G = (S, V_N, V_T, R)$ , où  $V_N$  est un ensemble de symboles, nommé *vocabulaire non terminal*,  $V_T$  est un ensemble de symboles nommé *vocabulaire terminal*,  $S$  le *symbole initial* et  $R$  un ensemble de *règles de réécriture*.

Une *grammaire catégorielle* assigne des catégories syntaxiques à chaque unité linguistique. Les catégories syntaxiques sont des types orientés, engendrés à partir de types de base et de deux opérateurs constructifs '/' et '\':

(i)  $N$  (syntagme nominal) et  $S$  (phrase) sont des types de base.

(ii) Si  $X$  et  $Y$  sont des types orientés alors  $X/Y$  et  $X \setminus Y$  sont des types orientés

Les deux règles d'application (avant et arrière) sont notées ainsi :

$$\frac{[X \setminus Y : u1] \quad [Y : u2]}{[X : (u1 \setminus u2)]} > \quad \frac{[Y : u1] \quad [X \setminus Y : u2]}{[X : (u2 \setminus u1)]} <$$

Les grammaires génératives privilégient la structure concaténée de la phrase tandis que les grammaires catégorielles privilégient la structure

<sup>9</sup> Une analyse comparative des trois grammaires catégorielles proposées par Adjukiewicz, Bar Hillel et Lambek est donnée dans Godard-Wendling (2002).

<sup>10</sup> Desclés et Biskri définissent la grammaire catégorielle combinatoire applicative. Le passage de la structure concaténée à la structure applicative se fait en rajoutant aux règles de la grammaire catégorielle les combinateurs  $C^*$  et  $B$  et  $S$  de Curry.

<sup>11</sup> Une analyse de la distinction entre type et catégorie est donnée dans Bourdeau (2002).

fonctionnelle (opérateur-opérande). Les deux types de grammaires sont définis dans un langage symbolique.

En ce qui concerne le langage symbolique, pour les grammaires génératives la syntaxe est représentée par : un symbole (S) deux ensembles des primitives ( $V_N$ ,  $V_T$ ) et un ensemble de règles (R). Leur sémantique est représentée par la description du fonctionnement de la grammaire.

Pour les grammaires catégorielles, la syntaxe est formée d'un ensemble de types (de base et composés) et de deux règles ( $>$ ,  $<$ ). La sémantique de ces grammaires est la transformation de la représentation concaténée de la phrase en représentation fonctionnelle.

La syntaxe et la sémantique formelles ont beaucoup aidé, d'une part, l'analyse des langues et, d'autre part, la construction des langages formels qui, à leur tour, ont contribué à la réalisation des logiciels d'analyse morphologique et syntaxique ou des logiciels pour la fouille sémantique des textes<sup>12</sup> ou de la traduction automatique<sup>13</sup>.

Les grammaires modélisent la structure de la phrase en lui associant une expression (représentation) logique. Une analyse sémantique de texte fondée sur la grammaire est plus efficace quand on cherche une fidélité (exactitude) élevée sur des textes courts contenant un langage complexe.

Enfin, nous pouvons formuler que le langage iconique est un langage qui utilise comme objets des représentations liées plutôt aux images tandis que le langage symbolique est un langage qui utilise d'autres symboles pour ces objets.

Nous pouvons affirmer aussi que, pour décrire le même contenu de pensée, entre le langage iconique et le langage symbolique, il y a un processus cognitif de transformation, de « traduction complexe » du type de la *compilation généralisée* (Abraham *et al.* 1990 ; Sauzay 2013) comme il est représenté dans la figure 9. Dans cette figure, le LI, LI<sub>1</sub>,...LI<sub>n</sub>, représentent des langages iconiques de plus en plus proches d'un symbolisme. Le passage de l'un à l'autre se fait par un type de compilation qui n'est pas autre chose qu'une traduction particulière. Par exemple, le passage d'une image au sens commun, à un schéma et, encore, à un graphe

---

<sup>12</sup> Les modèles d'annotation sémantique et filtrage sémantique des textes développés dans le cadre du laboratoire LALICC de l'université Paris 4 fondés sur la méthode de l'exploration contextuelle entre 1995 et 2008 ont été intégrés dans le logiciel EXCOM (Alrahabi Motasem, Desclés Jean-Pierre [2009], « EXCOM : Plate-forme d'annotation sémantique de textes multilingues », TALN 2009, Senlis, 24-26 juin 2009).

<sup>13</sup> Actuellement, pour la traduction automatique, il existe deux grands types de traducteurs automatiques : traduction à base de règles (donc implicitement, basée sur la grammaire) et traduction basée sur des algorithmes statistiques (T.A.S).

est la traduction entre  $LI_0$ ,  $LI_1$ ,  $LI_2$ . Le langage  $LI_0$  est le langage iconique le plus primitif, l'*ancrage* du concept représenté dans le réseau de concepts auquel il appartient. Ensuite, le passage entre  $LI_n$ , le dernier langage iconique, et  $LS_0$ , le premier langage symbolique se fait toujours par une traduction. La chaîne continue avec des langages symboliques de plus en plus affinés. Comme exemple, on peut considérer le langage symbolique de Frege comme étant  $LS_0$  et le langage symbolique de la logique moderne comme étant  $LS_n$ . En sens inverse, le processus cognitif se déroule pour le *décodage du sens*, dans un mélange de pensée heuristique et procédurale pour les niveaux symboliques et plutôt heuristique pour les niveaux iconiques.

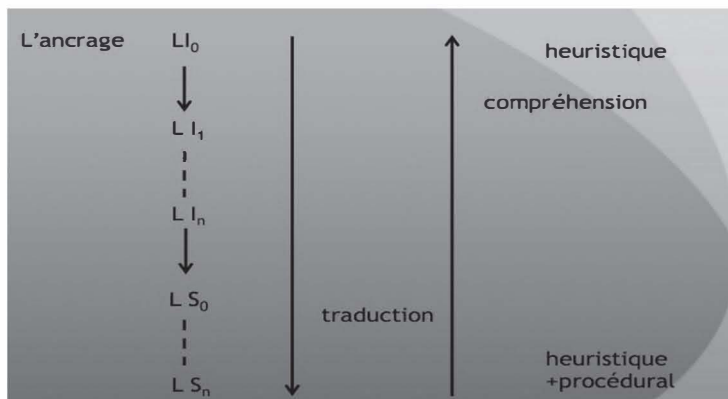


Figure 9. Langage iconique – langage symbolique

## Conclusion

En conclusion, nous postulons l'existence de deux types de langage pour le traitement (construction, acquisition et transmission) des connaissances : le langage iconique et le langage symbolique. Si l'on accepte la dichotomie *langage iconique – langage symbolique*, les traits caractéristiques de ces deux types de langage sont les suivants :

Pour le langage iconique :

- Ses primitives sont proches de l'image ;
- Son système de règles est proche des opérations de transformation de l'image ;
- L'ambiguïté dans l'interprétation peut être assez élevée.

Pour le langage symbolique :



- Ces primitives sont des symboles fixés a priori ;
- Son système de règles est plus riche que le système de transformations des images ;
- L'ambiguïté dans l'interprétation est réduite.

Dans le traitement des connaissances - construction, acquisition, transmissions - on opère avec les deux types. Le passage d'un type à l'autre se fait par une sorte de « compilation généralisée ». Le « traitement numérique » qu'on peut donner à certains problèmes dans tous les domaines des sciences se fait en utilisant des langages symboliques. On part d'un problème d'un certain domaine scientifique. On construit un modèle qui lui est associé. Ce modèle est une représentation soit mathématique (y compris, statistique), soit logique, soit appartenant à un autre domaine scientifique. Le modèle utilise nécessairement un langage symbolique propre. Ce langage symbolique construit des représentations des éléments cognitifs du problème modélisé qui sont « traduites » dans un langage de programmation exécuté par la machine selon un algorithme (procédure) spécifié par le modèle. Les résultats obtenus suite au « traitement numérique » suivent le chemin inverse. Ils sont retraduits dans le langage appartenant au domaine dont le problème est issu. Ce cheminement cognitif se trouve à la base de ce que nous appelons aujourd'hui « interdisciplinarité » entre l'informatique et un autre domaine, ou « traitement automatique » d'un problème appartenant à un certain domaine. Il représente la plus importante, la plus utile et, en même temps, la plus difficile partie du « numérique » où la créativité humaine intervient et le langage symbolique apporte sa contribution.

## Bibliographie

- Arnauld, Antoine et Pierre Nicole. 1964. *Logique de Port-Royal (1662)*, Presses de l'université de Lille
- Arnauld, Antoine et Pierre Nicole, 1992. *La Logique ou l'art de penser (1662)*. Paris, Gallimard. Introduction Louis Marin.
- Auroux, Silvain. 1993. *La logique des idées*. Paris : Vrin, Bellarmine.
- Abraham, Maryvonne, Jean-Pierre Desclés, David Piotrowski et Frédérique Segond. 1990. « Langage naturel et représentations cognitives : un problème d'architecture et de compilation ». Actes du 4<sup>e</sup> colloque de l'ARC, Progrès de la science cognitive : 290-303.
- Banks, David. 2005. *Introduction à la linguistique systémique de l'anglais*. Paris : L'Harmattan.



- Banks, David. 2011. « La Linguistique Systémique Fonctionnelle : une approche sémantique et sociale », Université de Brest. *Amadis*, no.9: 265–278.
- Biard, Joël. 1989. *Logique et théorie du signe au XIV<sup>e</sup>*. Paris: J. Vrin, Librairie Philosophique.
- Biskri, Ismail et Jean-Pierre Desclés. 1995. « Logique combinatoire et linguistique : grammaire catégorielle combinatoire applicative ». *Mathématique et sciences humaines*, Paris MSH –MSS, no. 132.
- Bourdeau, Michel. 2002. « La genèse des grammaires catégorielles et leur arrière-plan logico-philosophique : quelques remarques », *Langages*, volume 36, no. 148 : 13–27.
- Busquets, Joan. 2006. *Logique et langage : apport de la philosophie médiévale*. Bordeaux : Presses Universitaires de Bordeaux.
- Chomsky, Noam. 1965 *Aspects of the Theory of Syntax*, MIT Press.
- Desclés, Jean-Pierre. 1990. *Langages applicatifs, langues naturelles et cognition*. Paris : Hermès.
- Desclés, Jean-Pierre et Anca Pascu. 2011 « Logic of Determination of ●objects (LD●): How to Articulate “ Extension ” with “ Intension ” and “ ●objects ” with “ Concepts ” ». Springer, *Logica Universalis*, Volume 5, Issue 1: 75–85.
- Frege, Gottlob. 1971a. *Ecrits logiques et philosophiques, (1879–1925)*. Traduction de C. Imbert. Paris : Editions du Seuil.
- Frege, Gottlob. 1971b. *Les fondements de l'arithmétique (1893)*, En L'ordre philosophique, collection dirigée par P. Ricoeur et F. Wahl, traduction de C. Imbert. Paris : Editions du Seuil.
- Frege, Gottlob. 1967. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet (1893)*. University of California Press, edition, 1893. Translated and edited with an introduction by Montgomery Furth : Basic Laws of Arithmetic (exposition of the system).
- Godin, Christian. 2004. *Dictionnaire de philosophie*, Paris, Fayard, 1534 p.
- Husserl, Edmund. 1969. *Recherches logiques : Recherches pour la phénoménologie et la théorie de la connaissance*. Paris : Presses Universitaires de France, volume 2 Titre original *Logische Untersuchungen ; Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*, publié par M. Niemeyer à Halle en 1913, traduit par H. Elie, A. Kelkel et R. Scherer.
- Kleene, Stephen Cole. 1971. *Logique mathématique*. Paris : Armand Colin.
- Pascu, Anca Christine. 2001. « Logique de détermination des objets ». Thèse de doctorat. Paris : Université de Paris-Sorbonne.
- Peirce, Charles Sanders. 1995. *Le raisonnement et la logique des choses – Les conférences de Cambridge (1898)*. Introduction par Kenneth Laine

- Ketner et Hilary Putnam, traduction de l'anglais par Christiane Chauviré, Pierre Thibaud et Claudine Tiercelin. Paris : Les Editions du Cerf,
- Russell, Bertrand. 1989. *Ecrits de logique philosophique (1902-1918)*. Traduction J.-M. Roy, Paris : Presses Universitaires de France, Epiméthée.
- Russell, Bertrand. 1937. *The Principles of Mathematics*. Londres : G. Allen & Unwin.
- Sauzay, Benoît. 2013. « Le concept informatique de compilation généralisée dans les sciences cognitives (linguistique, logique et intelligence artificielle) : contributions aux rapports entre la logique combinatoire et les T[■] – algèbres ». Thèse de doctorat. Paris : Université de Paris-Sorbonne.
- Shaumyan, K. Sebastian. 1987. *A Semiotic Theory of Natural Language*, Bloomington : Indiana Univ. Press. Van Dalen, Dirk. 2013. *Logic and Structure*. Springer 1997, ré-edition.
- Van Dalen, Dirk. 2013. *Logic and Structure*, London/New York : Springer, 5ème éd.
- Vernant, Denis. 2009. « Russell et Whitehead ». Dans *A. N. Whitehead : Le procès de l'univers et des savoirs*, L'Art du Comprendre, no. 18 : 105–113.

## CHAPTER ELEVEN

# L'ÉVOLUTION PEDAGOGIQUE DE LA LANGUE DES LINGUISTES ALLEMANDS AU COURS DU XIX<sup>E</sup> SIECLE

## THE EVOLUTION OF THE DISCOURSE OF GERMAN LINGUISTS DURING THE 19<sup>TH</sup> CENTURY

JACQUES FRANÇOIS

### **Abstract**

This contribution aims to echo the double enterprise of David Banks, which consists in illustrating Systemic Functional Linguistics (SFL) while investigating the scientific discourse of the classical age in France. It focuses on three excerpts of works from 19th century German linguists. In the first half of the century, illustrated here by the writings of Wilhelm von Humboldt and Franz Bopp, linguistics has difficulty distancing itself from philology and develops within a limited circle of specialists. The linguistic discourse is directed at peers. However, by the end of the century, with the rapid expansion of universities in Germany following the country's unification in 1870, this discourse takes on an educational dimension in order to train language teachers by informing them of the state of science. This is illustrated in the excerpt from Meyer-Lübke. By distinguishing between different types of relationships between clauses, SFL is a convenient tool for comparatively assessing how the syntactic structure of these discourses evolved during the 19th century.

Le propos de cette contribution est double : David Banks est à la fois celui qui a diffusé en France la Linguistique Fonctionnelle Systémique (LFS) et celui qui l'a appliquée à un champ d'étude spécifique, l'analyse du discours scientifique dans la France de l'Ancien régime, à travers ses études sur le *Journal des Sçavans*. Je chercherai ici à faire écho à cette entreprise jumelée en examinant

- l'évolution de la langue des linguistes allemands, lorsque – essentiellement sous l'effet de l'unité allemande en 1870 – ils ont été amenés à former des quantités croissantes d'étudiants et
- la capacité de la LFS à décrire cette évolution.

Mais cette brève étude n'a d'autre prétention qu'illustrative : les trois exemples sélectionnés de la langue scientifique de Wilhelm von Humboldt, Franz Bopp et Wilhelm Meyer-Lübke sont symptomatiques, à mon sens, d'un « mode d'interlocution » variable : les deux premiers s'adressent à des pairs dans la première moitié du XIX<sup>e</sup> siècle, le troisième aux tout nouveaux étudiants en linguistique romane au tournant du XX<sup>e</sup> siècle<sup>1</sup>. Si Humboldt et Bopp avaient été en présence d'étudiants d'origines diverses et en nombre trop important pour suivre la progression de chacun d'eux, ils auraient peut-être adopté eux aussi un langage adapté à leurs auditeurs.

*De facto*, dans la seconde moitié du XIX<sup>e</sup> siècle l'enseignement universitaire est devenu en Allemagne une priorité nationale afin de développer le *Bildungsbürgertum* (la bourgeoisie cultivée) porteur d'un *Kulturbesitz* (patrimoine culturel, cf. Hülten Schmidt 2000) comme levier social en marge de l'aristocratie terrienne, ce qui va permettre l'épanouissement de certaines des plus prestigieuses universités d'Europe, particulièrement Leipzig et Berlin pour la linguistique, d'où l'adoption d'un discours désormais accessible à un plus grand nombre.

## 1. Introduction bipartite

Considérons en premier lieu l'environnement culturel et disciplinaire des fragments destinés à illustrer l'évolution du discours linguistique (§1.1) et l'outil que nous offre la LSF pour cet examen (§1.2).

---

<sup>1</sup> Pour une introduction développée à l'œuvre de ces trois linguistes, on pourra se reporter à François (2017).

### 1.1. Qui est linguiste en Allemagne au XIX<sup>e</sup> siècle ?

Le premier linguiste au sens étroit que nous donnons aujourd'hui à ce terme, c'est-à-dire le premier savant uniquement préoccupé de la mise en évidence de régularités dans une langue, à travers son histoire ou dans un éventail de langues, a été incontestablement<sup>2</sup> Franz Bopp, l'un des tout premiers universitaires à se consacrer exclusivement à l'étude comparative des langues indo-européennes, appelé à l'université de Berlin par son fondateur en 1809, Wilhelm von Humboldt. On ne lui connaît aucune édition philologique ni aucune étude consacrée aux mythologies (il laissait cela par exemple à son collègue Max Müller, l'éditeur des *Rigveda*, qui régnait à l'époque sur l'indologie du haut de sa chaire londonienne) et il se consacrait uniquement à la comparaison des langues anciennes qui présentaient des affinités structurales plus ou moins évidentes (cf. Bopp 1816).

Humboldt, de son côté, était un aristocrate exceptionnellement cultivé, tout comme son frère, le célèbre géographe Alexander von Humboldt. Il était fasciné par la variété des langues du monde, et dans l'esprit des Lumières, il était à la recherche de l'unité de l'esprit humain au-delà de cette diversité. Mais, inspiré par son maître et ami Goethe, il s'intéressait à une multitude de questions d'histoire, d'anthropologie et de philosophie en marge de sa carrière de diplomate qui allait le conduire en 1815 au Congrès de Vienne comme représentant de la Prusse.

Le profil de Wilhelm Meyer-Lübke nous est plus familier : professeur de romanistique à l'université de Vienne, il a repris dans le dernier quart du siècle l'ambition de Friedrich Diez ; éditer une grammaire comparée et un dictionnaire des étymologies romanes. Ces ouvrages massifs allaient représenter le nouvel état du savoir un demi-siècle après son maître, dans une discipline qui était aux avant-postes de la recherche sur les langues indo-européennes en raison de la reconstruction relativement aisée du latin vulgaire, comparée à celle des états primitifs des langues germaniques, slaves ou celtes.

Au début du XIX<sup>e</sup> siècle, il n'y a pas encore en Allemagne (et encore moins à l'étranger) de linguistes professionnels. Avec la fondation, d'un côté de la grammaire historico-comparative des langues indo-européennes et, d'un autre côté, des bases de l'ethnolinguistique (dans le prolongement du *Mithridates* de J. Ch. Adelung et J.S. Vater, 1806–1817), les deux orientations académiques majeures de la linguistique se mettent en place au

---

<sup>2</sup> C'est aussi l'opinion de Michel Foucault (*Les mots et les choses*, 1965, 145sq) qui voit en Bopp l'équivalent pour la linguistique, Ricardo pour l'économie et Cuvier pour la paléontologie.

cours du siècle. Jusqu'aux années 1820, c'est l'enseignement de la philologie classique qui domine, puis l'enseignement des philologies modernes (essentiellement des littératures médiévales allemande, anglaise, française et italienne) se met en place. À la même époque, la linguistique « générale » joue un rôle négligeable dans l'enseignement, défendue par quelques originaux comme Heyman Steinthal à Berlin ou Friedrich Müller à Vienne.

## **1.2. Les opérations « logico-sémantiques » de base (expansion, projection, élaboration, adjonction et cadrage) de la LSF et la relation syntaxique d'enchâssement**

Venons-en maintenant à l'outil permettant de comparer – entre syntaxe et sémantique – les trois fragments présentés dans les sections 2 à 4. Halliday & Matthiessen (2004, 377) introduisent deux types d'opérations qu'ils qualifient de « logico-sémantiques », l'EXPANSION et la PROJECTION, et trois sous-types de l'expansion (élaborante / *elaboration*, adjoignante / *extension* et cadrante<sup>3</sup> / *enhancement*) et deux types de projections (élocutive / *locution* et mentale / *idea*), cf. tableau 7(5), p. 380 (p.447 dans la 4<sup>e</sup> éd. de 2014, avec un développement) :

---

<sup>3</sup> « Cadrant » n'est pas une traduction transparente de « enhancing ». Mais la « enhancing expansion » concerne les subordonnées circonstancielles, dont le rôle consiste métaphoriquement à cadrer l'événement ou la situation par la mention d'un éventail de circonstances (localisation, cause, finalité, etc.). En outre, le cadrage peut s'entendre sur le plan syntaxique comme sur le plan sémantique, ce qui est conforme à l'intention des auteurs.



Tableau 1 : Les types logico-sémantiques d'expansions et de projections selon Halliday & Matthiessen (2005: 380)

		(i) paratactic	(ii) hypotactic
(1) expansion	(a) elaboration	1 John didn't wait; -2 he ran away. 'apposition'	α John ran away, -β which surprised everyone. 'non-defining relative'
	(b) extension	1 John ran away, +2 and Fred stayed behind. 'co-ordination'	α John ran away, β whereas Fred stayed behind.
	(c) enhancement	1 John was scared, :2 so he ran away.	α John ran away, ×β because he was scared. 'adverbial clause'
(2) projection	(a) locution	1 John said: "2 'I'm running away" 'direct speech'	α John said "β he was running away. 'indirect speech'
	(b) idea	1 John thought to himself: '2 'I'll run away'	α John thought 'β he would run away.

Les propositions relatives et participiales relèvent de l'expansion élaborante (' : *elaborating expansion*) et sont spécifiables comme (a) finies vs non-finies et (b) restrictives (*defining*) vs descriptives (*non-defining*), cf. Tableau 7(12), p.427 (p.492 dans la 4<sup>e</sup> éd. de 2014) :

Tableau 2 : Les types d'enchâssement (changement de rang syntaxique) selon Halliday & Matthiessen (2004: 427)

Function	Class	Innominal group	In adverbial group
Postmodifier	clause: finite	the house [(that Jack built)]	sooner [(than we had expected)]
	clause: non-finite	the house [(being built by Jack)]	sooner [(than expected)]
	phrase	the house [by the bridge]	sooner [than the rest of us]
Head	clause: finite	[[what Jack built]]	—
	clause: non-finite	[[for Jack to build a house]]	—
	phrase	[by the bridge]	—

Les propositions coordonnées ou reliées par un connecteur ou par une conjonction subordonnante de simultanéité relèvent de l'expansion adjoignante ('+' : *extending expansion*) et les propositions circonstancielles relèvent de l'expansion cadrante ('x' : *enhancing expansion*).

## 2. La linguistique philosophique et la langue de W. Von Humboldt (1823)

Humboldt avait une capacité exceptionnelle, celle de faire de la philosophie du langage à partir d'exemples concrets empruntés à une multitude de langues. L'extrait suivant de sa conférence sur « Les verbes dans les langues américaines » en 1823 à l'Académie de Berlin en est un exemple. La partie étudiée plus bas est la troisième phrase en italiques.

<p>Der Begriff des Seyn's, welcher die Grundlage, und das Wesen des Verbum ausmacht, kann angedeutet seyn</p> <ul style="list-style-type: none"> <li>• entweder als für sich bestehen,</li> <li>• oder als der Verbalform körperlich, als Hülfverbum, einverleibt,</li> <li>• oder als in ihr nur der Idee nach enthalten.</li> </ul> <p>Nach diesen drei Abteilungen werden sich die Verschiedenheiten der verglichenen Sprachen am richtigsten stellen lassen. <i>Nur darf man nicht vergessen, daß es wohl keine Sprache geben mag, in welcher nicht die erste dieser Methoden mit einer der letzten zugleich gebraucht werden sollte, und daß in denjenigen Sprachen, welche ein Verbum substantivum und in ihrer Conjugation Formen mit und ohne Hülfverba besitzen, alle diese drei Fälle vorkommen.</i></p>	<p>Le concept de l'être, qui constitue le fondement et l'essence du verbe peut-être indiqué</p> <ul style="list-style-type: none"> <li>• soit comme existant par lui-même</li> <li>• soit comme incorporé à titre de verbe auxiliaire à la forme verbale,</li> <li>• soit encore comme contenu idéellement dans celle-ci.</li> </ul> <p>Après ces trois sections, il sera possible d'identifier le plus correctement les différences entre les langues en contraste. <i>Il ne faut toutefois pas oublier qu'il n'y a probablement aucune langue dans laquelle le premier de ces procédés ne serait pas employé simultanément à l'un des deux autres et que dans les langues qui ont un verbum substantivum et qui possèdent dans leur conjugaison des formes avec et sans auxiliaires, ces trois procédés sont tous présents.</i></p>
--	---

Le tableau 3 offre une pré-analyse de ce fragment croisant la perspective syntagmatique (« l'ordre linéaire » de Tesnière 1959) et la perspective dépendancielle (son « ordre structural »). Du point de vue syntagmatique, le fragment se compose de cinq propositions finies et du point de vue dépendanciel, celles-ci se répartissent sur trois niveaux N1-N3. À chaque proposition est donc associable une coordonnée linéaire & dépendancielle : (Pr)1-N1, Pr2 : 2-N2, 3-N3, 4-N2, 5-N3. Cette pré-analyse ne spécifie pas

encore le statut morphosyntaxique de ces propositions ni leur classement logico-sémantique.

Tableau 3 : Pré-analyse du fragment de Humboldt (original et traduction isomorphe) selon l'ordre linéaire et l'ordre structural (cf. Tesnière 1959)

Pr	N1	N2	N3		N1	N2	N3
1	Nur darf man nicht vergessen,				Il ne faut toutefois pas oublier		
2	↘	daß es wohl keine Sprache geben mag,			↘	qu'il n'y a probablement aucune langue	
3		↘	in welcher nicht die erste dieser Methoden mit einer der letzten zugleich gebraucht werden sollte,		↘	dans laquelle le premier de ces procédés ne serait pas employé simultanément à l'un des deux autres	
4	→	und daß in denjenigen Sprachen, [...↘...] alle diese drei Fälle vorkommen.			→	et que dans les langues [...↘...] ces trois procédés sont tous présents.	
5		↘	welche ein Verbum substantivum und in ihrer Conjugation Formen mit und ohne Hülfverba besitzen		↘	qui ont un <i>verbum substantivum</i> et dans leur conjugaison des formes avec et sans auxiliaires.	

Pour l'analyse proprement dite, étant donné que j'ai proposé une traduction isomorphe, je me contente de raisonner à partir de la traduction en français, ce raisonnement étant applicable aussi bien à l'original en allemand. Du point de vue morphosyntaxique, les cinq propositions sont de trois types, une principale (1), deux conjonctives-complétives coordonnées (2, 4) et deux relatives restrictives (3, 5). Du point de vue systémique, la principale introduit deux projections mentales successives (notation  $\alpha - \beta$ ,  $+\beta$ ) et chacune de celles-ci comporte un référent nominal spécifié par une relative restrictive (« defining relative clause ») : *aucune langue dans laquelle..., dans les langues qui ont...*

Ce fragment est donc syntaxiquement transparent, avec un effet de balancement produit par le parallélisme des deux relatives introduites par les deux conjonctives-complétives.

Tableau 4 : Analyse morphosyntaxique et systémique de la traduction française isomorphe du fragment de Humboldt

Pr	type morpho-syntaxique	structure systémique	Français niveaux syntaxiques		
			N1	N2	N3
1	principale	α	Il ne faut toutefois pas oublier		
2	conjonctive-complétive	'β (projection mentale)	↘	qu'il n'y a probablement aucune langue	
3	relative (restrictive) enchâssée	[[clause : defining relative]]		↘ ↙	dans laquelle le premier de ces procédés ne serait pas employé simultanément à l'un des deux autres
4	conjonctive-complétive coordonnée	+β (extension à une 2 <sup>e</sup> projection mentale)			et que dans les langues [... ↘ ...] ces trois procédés sont tous présents.
5	relative (restrictive) enchâssée	[[clause : defining relative]]		↘	qui ont un <i>verbum substantivum</i> et dans leur conjugaison des formes avec et sans auxiliaires.

### 3. La grammaire historico-comparative et la langue de Franz Bopp (1836)

Le second fragment, emprunté à un ouvrage de Franz Bopp daté de 1836, a une structure syntaxique beaucoup plus complexe, caractérisée par une articulation périodique en protase et apodose :

<p>(p.VIII) Obwohl ich in meiner Kritik über Grimm's vortreffliche Grammatik nicht die Absicht hatte, vorzüglich in phonetische Erörterungen einzugehen, sondern vielmehr in dem Gange, den mehr zufällig als vorherbestimmt meine Untersuchung nahm, aus Mangel an Raum, Grimm's umfassende und scharfsinnige Lautlehre unbesprochen bleiben mußte :</p>	<p>PROTASE</p>	<p>Même si dans ma critique de l'excellente grammaire de Grimm je n'ai pas eu l'intention d'entrer particulièrement dans des discussions phonétiques, et si, bien plutôt, dans le cours que mon étude a pris plus par hasard que par résolution, la phonétique vaste et lumineuse de Grimm n'a pas pu être prise en compte, faute de place,</p>
<p>so drehten sich doch meine grammatischen und sprachvergleichenden Beobachtungen hauptsächlich um den Vocal, dieses feinere, höchst wandelbare Element des Sprachkörpers, das bei allen grammatischen Bestimmungen mit in Betracht kommt, in seinen Metamorphosen aber nicht so leicht wieder erkennbar ist, als wenn etwa ein Consonante von der Stufe der Tenuis zu jener der Aspirata oder von da zur Media</p>	<p>APODOSE</p>	<p>mes observations grammaticales et comparatives ont cependant tourné essentiellement autour de la voyelle, cet élément subtil et extrêmement variable, qui entre en ligne de compte dans tous les inventaires grammaticaux, mais qu'il n'est pas aussi facile de reconnaître à travers ses métamorphoses que lorsque par exemple une consonne se présente comme fléchie du degré des Tenuis à celui des aspirées et de là jusqu'aux Media.</p>

Le tableau 5 montre qu'au niveau de la pré-analyse, la phrase s'organise circulairement autour de la principale (Pr5-N1), précédée de deux subordonnées coordonnées de niveau 2 (Pr1, Pr3), la première étant complétée par une proposition infinitive (...*je n'ai pas eu l'intention d'entrer...*) et la seconde comportant un référent nominal spécifié par une relative (...*le cours que ... mon étude a pris*), et suivie de deux relatives (Pr6, Pr7) spécifiant un même référent nominal (*cet élément ... qui entre en ligne de compte ... mais qu'il n'est pas aussi facile...*), la seconde introduisant elle-même une subordonnée du niveau 3 (*que lorsque ...* : Pr7).

Tableau 5 : Pré-analyse du fragment de Bopp (original et traduction)

Pr	N1	N2	N3	N1	N2	N3
1		●	wohl ich in meiner Kritik über Grimm's vortreffliche Grammatik nicht die Absicht hatte,			Même si dans ma critique de l'excellente grammaire de Grimm je n'ai pas eu l'intention
2		↘	vorzüglich in phonetische Erörterungen einzugehen,		↘	d'entrer particulièrement
3		↙	sondern vielmehr in dem Gange [... ↘ ...] aus Mangel an Raum, Grimm's unfassende und scharfsinnige Lautlehre unbesprochen bleiben mußte		↙	dans des discussions phonétiques,
4		↘	den mehr zufällig als vorherbestimmt meine Untersuchung nahm		↘	et si, bien plutôt, dans le cours [... ↘ ...] la phonétique vaste et humineuse de Grimm n'a pas pu être prise en compte, faute de place,
5			so drehten sich doch meine grammatischen und sprachvergleichenden Beobachtungen hauptsächlich um den Vocal, dieses feinere, höchst wandelbare Element des Sprachkörpers,			que mon étude a pris plus par hasard que par résolution,
6		↘	das bei allen grammatischen Bestimmungen mit in Betracht kommt,		↘	mes observations grammaticales et comparatives ont cependant tourné essentiellement autour de la voyelle, cet élément subtil et extrêmement variable de l'organe linguistique,
7		↘	in seinen Metamorphosen aber nicht so leicht wieder erkennbar ist,		↘	qui entre en ligne de compte dans tous les inventaires grammaticaux,
8		↘	als wenn etwa ein Consonante von der Stufe der Tenuis zu jener der Aspirata oder von da zur Media herabgesunken erscheint.		↘	mais qu'il n'est pas aussi facile de reconnaître à travers ses métamorphoses,
					↘	que lorsque par exemple une consonne se présente comme fléchie du degré des Tenuis à celui des aspirées et de là jusqu'aux Media.

Du point de vue fonctionnel (cf. tableau 6, les propositions 1 et 3 sont des concessives préposées véhiculant une fonction d'expansion « cadrante » (*enhancing expansion*) notée  $x\beta$  (par rapport à  $\alpha/Pr5$ ). L'infinitive (Pr2) spécifie *l'intention* manifestée en Pr1, à savoir celle *d'entrer dans des discussions phonétiques*, avec le statut fonctionnel d'un « enchâssement



cadrant » (*enhancing embedding*, cf. Halliday & Matthiessen 2004, 442)<sup>4</sup> noté  $x \gamma$  (par rapport à  $\beta$ /Pr1). La seconde concessive (Pr3) est coordonnée à la première (Pr1), noté  $+ x \beta$  (par rapport à  $\alpha$ /Pr5). En outre, comme nous l'avons déjà vu dans le tableau précédent, Pr4 a le statut de relative restrictive spécifiant le référent nominal *le cours* (*que...*). L'apodose est constituée de son côté de deux relatives restrictives spécifiant le référent nominal *cet élément...* (*qui ... mais qu'il n'est pas aussi facile...*), la seconde introduisant une comparative hypothétique de niveau 3 (*aussi facile... que lorsque ... une consonne se présente ...*).

● On retrouve donc ici le balancement du fragment de Humboldt avec les deux concessives préposées introduisant l'une une complétive infinitive (Pr2), l'autre une relative restrictive (Pr4), mais confiné dans la protase et subordonné à un balancement d'ordre supérieur, celui entre les quatre propositions de la protase et les trois de l'apodose.

---

<sup>4</sup> Halliday & Matthiessen (2004) donnent comme exemple équivalent *the purpose [of raising funds]*.

Tableau 6 : Analyse morphosyntaxique et systémique de la traduction française isomorphe du fragment de Bopp

N°	structure syntaxique	structure systémique	N1 N2 N3
1	concessive préposée	x β (▶α/5)	Même si dans ma critique de l'excellente grammaire de Grimm je n'ai pas eu l'intention
2	compl. infinitive	x γ (▶β/1)	↘ d'entrer particulièrement dans des discussions phonétiques, ↙
3	sub. concessive coordonnée	+ x β (▶α/5)	et si, bien plutôt, dans le cours [... ↘...] la phonétique vaste et lumineuse de Grimm n'a pas pu être prise en compte, faute de place
4	relative (restrictive)	↔ [[clause: defining rel.]]	↘ que mon étude a pris plus par hasard que par résolution,
5	principale	α	mes observations grammaticales et comparatives ont cependant tourné essentiellement autour de la voyelle, cet élément subtil et extrêmement variable de l'organe linguistique,
6	relative (restrictive)	↔ [[clause: defining relative]]	↘ qui entre en ligne de compte dans tous les inventaires grammaticaux,
7	relative (restrictive) coordonnée	↔ [[clause: defining relative]]	↓ mais qu'il n'est pas aussi facile de reconnaître à travers ses métamorphoses,
8	comparative-hypo-thétique	x β (▶α/5)	↘ que lorsque par exemple une consonne se présente comme fléchie du degré des Tenuis à celui des aspirées et de là jusqu'aux Media.

#### 4. La linguistique pédagogique et la langue de W. Meyer-Lübke (1909)

Le troisième fragment diffère des deux premiers à un double point de vue. En premier lieu, c'est un paragraphe constitué de quatre phrases et non d'une seule phrase. Ensuite il est emprunté à un manuel pour étudiants s'engageant dans un cycle de romanistique, discipline prestigieuse et ambitieuse des universités allemandes depuis le deuxième quart du XIX<sup>e</sup> siècle, qui impose la connaissance irréprochable d'au moins deux langues romanes et celle au moins superficielle de toutes les autres, et cela d'un point de vue descriptif, philologique, culturel et linguistique. Ce fragment

constitue l'entrée en matière du chapitre sur les caractères communs au groupe des langues romanes.

[Ph1] Unter romanischen Sprachen versteht man die aus dem Lateinischen hervorgegangenen, in ihrem Wortschatze, ihrer Syntax und ihrem Formenbau durchaus lateinisches Gepräge tragenden Sprachen. [Ph2] Dabei ist es vor allem der Formenbau, der den Charakter deutlich zeigt. [Ph3] Die neuromanische Syntax nämlich erweist sich in sehr vielen Dingen der neuhochdeutschen oder der neugriechischen näher verwandt als der altlateinischen, so daß, wenn man gezwungen wäre, ohne Kenntnis der historischen Verhältnisse die Beziehung dieser verschiedenen Sprachen aus der Syntax zu bestimmen, das Ergebnis sehr leicht zu ungunsten des Lateinischen ausfallen könnte. [Ph4] Und was den Wortschatz betrifft, so sind z.B. im Rumänischen die nichtlateinischen Elemente so zahlreich, daß sie, wenigstens numerisch, das Übergewicht haben und dennoch ist das Rumänische eine romanische Sprache, umgekehrt umfaßt im Albanesischen das lateinisch-romanische Sprachgut über 20%, das altertümlich kaum 10% der Stammwörter, und trotzdem gehört das Albanesische nicht mehr zum Romanischen.

[Ph1] On entend par « langues romanes » les langues issues du latin véhiculant un format typiquement latin dans leur lexique, leur syntaxe et leur morphologie. [Ph2] Ce faisant, c'est surtout la morphologie qui manifeste clairement ce caractère. [Ph3] La syntaxe néoromane se révèle en effet plus proche à de nombreux points de vue de celle du haut-allemand moderne ou du grec moderne que de celle de l'ancien latin, si bien que, si l'on était forcé de déterminer à partir de la syntaxe les relations entre ces différentes langues sans connaissance des faits historiques, le résultat pourrait très facilement devenir défavorable au latin. [Ph4] Et pour ce qui concerne le lexique, en roumain par exemple les éléments étrangers au latin sont si nombreux que, sur le plan numérique à tout le moins, ils sont majoritaires, et cependant le roumain est une langue romane ; inversement en albanais le patrimoine lexical roman couvre plus de 20% des mots-racines contre seulement 10% pour ceux qui sont originels, et pourtant l'albanais est extérieur à l'espace roman.

La méthode analytique étant bien rodée, je sauterai ici la pré-analyse, mais je présenterai en parallèle l'original et la traduction isomorphe. La première

phrase se compose d'une principale comportant un référent nominal (*les langues*) spécifié en allemand par une longue participiale interposée<sup>5</sup> (avec le participe en queue de proposition et de niveau N2) et en français par une participiale postposée de niveau N2, l'une et l'autre équivalent fonctionnellement à une relative restrictive ( $\approx$  *celles de langues qui sont issues du latin et qui véhiculent un format typiquement latin...*).

Tableau 7.1 : 1<sup>re</sup> phrase du fragment de Meyer-Lübke (originale et traduction isomorphe)

Ph	Pr	allemand	struct. syst.	français
	P1	Unter romanischen Sprachen versteht man die <...↓...> Sprachen.		● On entend par "langues romanes" les langues
Ph1	P2	N2[aus dem Lateinischen hervorgegangenen, in ihrem Wortschatze, ihrer Syntax und ihrem Formenbau durchaus lateinisches Gepräge tragenden]N2	enchâssement [[clause: defining relative]]	N2[issues du latin et véhiculant un format typiquement latin dans leur lexique, leur syntaxe et leur morphologie]N2.

La phrase 2 a une composition analogue à une relative restrictive jouant le rôle de thème et la principale celui de rhème ( $\approx$  *Ce qui manifeste clairement ce caractère, c'est la morphologie*).

<sup>5</sup> Il est à noter que pour les germanophones lettrés, le traitement à l'écrit des subordonnées interposées (habituellement appelées « proposition qualificative » dans les grammaires françaises de l'allemand et « épithète participiale étendue », *erweitertes Partizipialattribut*, dans les grammaires allemandes) est facilité en général par le repérage d'une suite de deux morphèmes qui ne peuvent pas entretenir une relation dépendancielle directe, ex. ici *die*<Art.Def.PI> *aus*<Prép>. Seuls les cas de propositions participiales (ou syntagmes adjectivaux) dénué(e)s de déterminant peuvent entraver la lecture, ex.  $\emptyset$ <Det> *aus dem Stegreif gegebene Antworten* (des réponses fournies au déboté).

Tableau 7.2 : 2<sup>e</sup> phrase du fragment de Meyer-Lübke

P1	Dabei ist es vor allem der Formenbau,		Ce faisant, c'est surtout la morphologie
Ph2	N2[der den Charakter	enchâssement	N2[qui manifeste clairement ce
P2	deutlich zeigt]N2.	[[clause: defining relative]]	caractère]N2.

La phrase 3 est plus complexe, avec une principale P1-N1 (α) introduisant une subordonnée consécutive P2-N2 (notée + β pour une expansion adjoignante)<sup>6</sup>, laquelle introduit à son tour une subordonnée hypothétique P3-N3 complétée par une infinitive enchâssée (all. *wenn man gezwungen wäre, ... zu bestimmen* / fr. *si l'on était forcé de déterminer...*).

---

<sup>6</sup> Le statut d'expansion adjoignante (*extending expansion*) est plus pertinent que celui d'expansion cadrante (*enhancing expansion*) car les subordonnées consécutives introduites par all. *so dass* / fr. *si bien que* ne sont pas déplaçables en tête de phrase, contrairement aux propositions causales ou concessives, ce qui suggère un statut intermédiaire entre subordination et coordination.

Tableau 7.3 : 3<sup>e</sup> phrase du fragment de Meyer-Lübke

Ph 3	P 1	Die <b>neuro</b> manische Syntax nämlich erweist sich in sehr vielen <b>Dingen der</b> neuhochdeutschen <b>oder</b> der neugriechischen näher verwandt als <b>der</b> altlateinischen,	<b>α</b>	La syntaxe <b>neuro</b> mane se révèle en effet plus <b>proche à</b> <b>de nombreux points de vue</b> <b>de celle du haut-allemand</b> <b>moderne ou du grec</b> <b>moderne que de celle de</b> l'ancien latin,
	P 2	N2[ <b>so daß</b> <...↓...> <b>das Ergebnis sehr</b> <b>leicht zu ungunsten</b> <b>des Lateinischen</b> <b>ausfallen könnte]</b>	+ β	N2[si bien <b>que</b> <...↓...> le résultat <b>pourrait très</b> <b>facilement devenir</b> <b>défavorable au latin]</b> N2
	P 3	N3[ <b>wenn man</b> <b>gezwungen wäre,</b>	x γ	N3[ <b>si l'on était</b> <b>forcé de</b>
	P 4	N4[ <b>ohne</b> <b>Kenntnis der</b> <b>historischen</b> <b>Verhältnisse</b> <b>die Beziehung</b> <b>dieser</b> <b>verschiedenen</b> <b>Sprachen aus</b> <b>der Syntax zu</b> <b>bestimmen]</b> N4]N 3	enchâs- sement [[[infini- tive]]	N4[ <b>déterminer à</b> <b>partir de la</b> <b>syntaxe les</b> <b>relations entre</b> <b>ces différentes</b> <b>langues sans</b> <b>connaissance</b> <b>des faits</b> <b>historiques]</b> N4]N 3

Quant à la phrase 4, c'est la plus riche du point de vue de l'ordre linéaire (6 propositions), mais pas la plus complexe du point de vue de l'ordre structural. Elle débute par une subordonnée relative indéfinie à fonction cadrante P1-N2, notée 1 x β (le thème est limité au vocabulaire) suivie d'une première principale (P2-N1) qui introduit une subordonnée corrélatrice (notée 1 α). Les trois propositions suivantes (P4-P6) sont des indépendantes connectées par all. *umgekehrt* / fr. *inversement* (P4-P5) et all. *und trotzdem* / fr. *et pourtant* (P5-P6). Cette dernière phrase est donc beaucoup moins hiérarchisée que la précédente (deux niveaux de dépendance en Ph4 contre quatre en Ph3).



Tableau 7.4 : 4<sup>e</sup> phrase du fragment de Meyer-Lübke

	Und		Et
P1	N <sub>2</sub> [was den Wortschatz betrifft] <sub>N<sub>2</sub></sub> ,	1 x β	N <sub>2</sub> [pour ce qui concerne le lexicale] <sub>N<sub>2</sub></sub> ,
P2	N <sub>1</sub> [so sind z.B. im Rumänischen die nichtlateinischen Elemente so zahlreich] <sub>N<sub>1</sub></sub> ,	1 α	N <sub>1</sub> [en roumain par exemple les éléments étrangers au latin sont si nombreux] <sub>N<sub>1</sub></sub>
P3	N <sub>2</sub> [daß sie, wenigstens numerisch, das Übergewicht haben] <sub>N<sub>2</sub></sub>	+ β	N <sub>2</sub> [que, sur le plan numérique à tout le moins, ils sont majoritaires] <sub>N<sub>2</sub></sub> ,
Ph4			
P4	N <sub>1</sub> [und dennoch ist das Rumänische eine romanische Sprache] <sub>N<sub>1</sub></sub> ,	+ 2	N <sub>1</sub> [et cependant le roumain est une langue romane] <sub>N<sub>1</sub></sub> ;
P5	N <sub>1</sub> [umgekehrt umfaßt im Albanesischen das lateinisch-romanische Sprachgut über 20%, das altererbte kaum 10% der Stammwörter] <sub>N<sub>1</sub></sub> ,	+ 3	N <sub>1</sub> [inversement en albanais le patrimoine lexical roman couvre plus de 20% des mots-racines contre seulement 10% pour ceux qui sont originels] <sub>N<sub>1</sub></sub> ,
P6	N <sub>1</sub> [und trotzdem gehört das Albanesische nicht mehr zum romanischen] <sub>N<sub>1</sub></sub> .	+ 4	N <sub>1</sub> [et pourtant l'albanais est extérieur à l'espace roman] <sub>N<sub>1</sub></sub> .

Sur la base de ces trois analyses, nous pouvons construire un tableau comparatif<sup>7</sup> :

<sup>7</sup> Ratio N1/N+ : proportion de propositions de rang 1 (principales) par rapport au total des propositions de la phrase.

Tableau 8 : Comparaison des structures syntaxiques des trois fragments<sup>8</sup>

HUMBOLDT					
N1	N2	N3	NbPr/Ph	ratio N1/N+	
1	2 post	2 post	5	0,2	

BOPP					
N1	N2	N3	NbPr/Ph	ratio N1/N+	
1	2 pré 2 post	3 post	8	0,13	

MEYER-LÜBKE						
	N1	N2	N3	N4	NbPr/Ph	ratio N1/N+
Ph1	1	1 inter			2	0,5
Ph2	1	1 post			2	0,5
Ph3	1	1 post	1 pré	1 post	4	0,25
Ph4	4	1 pré. 1 post			6	0,67
Total	7	3 post 1 pré 1 inter	1 pré	1 post	14/4 = 3,5	0,5

Le fragment de Humboldt (1823) se compose d'une seule phrase constituée de cinq propositions, une principale, deux subordonnées postposées (de projection mentale) du 2<sup>e</sup> niveau et deux relatives restrictives du 3<sup>e</sup> niveau. Le ratio N1/N+ est bas (0,12), la phrase est donc fortement hypotaxique. Aucune subordonnée n'étant préposée, la structure globale est relativement élémentaire.

Le fragment de Bopp (1836) se compose lui aussi d'une seule phrase constituée de huit propositions, à savoir :

- une principale,
- dans la protase deux subordonnées (d'expansion cadrante) préposées du 2<sup>e</sup> niveau introduisant une complétive infinitive et une relative restrictive du 3<sup>e</sup> niveau,
- et dans l'apodose trois subordonnées postposées (deux relatives restrictives du 2<sup>e</sup> niveau et une comparative hypothétique du 3<sup>e</sup> niveau).

<sup>8</sup> Les subordonnées préposées imposant l'attente de la proposition régissante sont surlignées en gris.

Le ratio N1/N+ est encore plus faible (0,13), c'est-à-dire que la phrase est encore plus hypotaxique que celle de Humboldt, ce qui est lié à la répartition des subordonnées entre la protase et l'apodose. De ce fait, la structure de ce fragment est beaucoup plus complexe que celle du fragment de Humboldt, en raison d'une part du nombre de subordonnées et d'autre part du statut syntagmatique préposé des propositions 1 et 3 constituant la protase.

Quant au fragment de Meyer-Lübke (1909), il se compose de quatre phrases constituées de deux à six propositions pour une moyenne de 3,5 propositions par phrase et avec un ratio N1/N+ bien supérieur (0,5). La proportion de propositions préposées ou interposées est faible (une de chaque), et surtout la structure de la phrase 4 est faiblement hiérarchisée puisqu'elle ne comporte que deux subordonnées du 2<sup>e</sup> niveau contre quatre principales. En dépit de la présence d'une subordonnée du 4<sup>e</sup> niveau et d'une subordonnée préposée au 3<sup>e</sup> niveau, la structure est relativement transparente, ce qui est le trait distinctif d'un discours pédagogique.

Bien entendu, l'inventaire du nombre, de la hiérarchisation et de la nature morphosyntaxique et logico-sémantique des propositions ne suffit pas à différencier un discours scientifique destiné aux pairs d'un discours destiné aux étudiants. La longueur des propositions, en particulier de celles qui sont préposées ou interposées demanderait à être prise en compte et le degré de sophistication du vocabulaire est évidemment déterminant. Cependant la complexité syntaxique est un facteur crucial, car elle détermine la fluidité de la lecture et, de ce point de vue, il est sûr que le fragment de Bopp est celui qui demande au lecteur l'attention la plus vive<sup>9</sup>.

---

<sup>9</sup> Merci à Chris Gledhill pour sa suggestion d'inclure dans la procédure d'évaluation comparative le calcul de la « densité lexicale » de M. Halliday consistant à mesurer le rapport entre le nombre de mots lexicaux et le nombre de propositions. Ce calcul pose des problèmes de classement : p.ex. quels sont les verbes et les adverbes « lexicaux » par rapport aux verbes et adverbes « grammaticaux » ? En considérant que les auxiliaires et les adverbes à valeur discursive sont de nature grammaticale, j'obtiens pour Humboldt un ratio de 3,6 (18 lexèmes pour 5 propositions), pour Bopp un ratio de 6,0 (48 lexèmes pour 8 propositions) et pour Meyer-Lübke un ratio de 3,7 (22 lexèmes pour 6 propositions/phrases). Le bilan est donc comparable quant à la spécificité du discours de Bopp, beaucoup plus riche en lexèmes par proposition, mais il ne rend pas compte de la différence de complexité syntaxique établie par le ratio N1/N>1 entre le nombre de propositions de rang 1 (principales) et de rang>1 (subordonnées) entre le discours de Meyer-Lübke (7/14 = 0,5) et celui de Humboldt (1/5 = 0,2). Et surtout LA DENSITE LEXICALE NE DIT RIEN DE LA DISPOSITION PREPOSEE, INTERPOSEE OU POSTPOSEE DES SUBORDONNEES, les deux premières représentant une charge importante (surtout si la subordonnée est longue) pour la mémoire de travail.

## 5. Conclusion : une langue adaptée à l'ouverture de l'université wilhelminienne

À partir des données précieuses de Storost (2001) sur la création des chaires de philologie moderne dans l'espace germanophone au XIX<sup>e</sup> siècle, on constate concernant par exemple la romanistique (cf. François 2017, chap.1), que cinq chaires de philologie romane sont ouvertes entre 1801 et 1822 (Marbourg, Halle, Bâle, Berlin et Bonn). Ces chaires jouent un rôle pionnier, mais certaines ne deviennent pérennes (après le départ de leur premier destinataire) qu'assez tard (en 1846 pour Bonn, 1867 pour Bâle ou 1896 pour Marbourg). En effet, aucune nouvelle chaire n'est instituée pendant trente ans (de 1823 à 1851). C'est le début de l'unité allemande à partir du 3<sup>e</sup> quart du XIX<sup>e</sup> siècle, et son institution définitive en 1870, qui vont entraîner une démultiplication des chaires de philologie romane (5 entre 1862 et 1869, 18 entre 1871 et 1898), et il en est de même pour les autres philologies modernes.

C'est ainsi que l'université wilhelminienne voit affluer un nombre croissant d'étudiants en général et dans les nouvelles philologies modernes en particulier. Simultanément le statut disciplinaire des linguistes évolue. Dans la première moitié du siècle, on distinguait les « maîtres de langue » qui n'avaient qu'un statut subalterne et les professeurs de philologie qui représentaient le savoir, obligatoirement historique et comparatif, sur un groupe de langues. Avec la reconnaissance croissante des acquis de la grammaire historico-comparative, la place de la linguistique s'accroît et entraîne par ailleurs un appel d'air pour une linguistique générale véhiculée par des spécialistes de langues « exotiques » comme le sinisant Georg von der Gabelentz à Leipzig puis Berlin, des ethnolinguistes comme Friedrich Müller à Vienne ou des psycholinguistes comme Heyman Steinthal à Berlin.

Là où les étudiants affluent, la rédaction de manuels s'impose. Et certains linguistes vont développer un réel talent en ce sens, en particulier August Schleicher (1859) pour les langues germaniques, Herman Paul (1880) pour les principes de la linguistique historique, Georg von der Gabelentz (1881) pour les questions de méthode en linguistique historique et générale, Berthold Delbrück pour l'indo-européen (1884) et les langues de l'Inde, ou Wilhelm Meyer-Lübke (1909) pour les langues romanes.

## Références

- ADELUNG, Johann Christoph & Johann Severin VATER. 1806–1817. *Mithridates, oder allgemeine Sprachenkunde*, 4 vol. [Mithridate, ou science générale des langues]. Berlin

- AUROUX, Sylvain, E.F.K. KOERNER, Hans-Joseph NIEDEREHE & Kees VERSTEEGH (eds.). 2000. *History of The Language Sciences / Geschichte der Sprachwissenschaften / Histoire des Sciences du Langage: An International Handbook [...]*. [vol.1]. Berlin : De Gruyter.
- BOPP, Franz. 1816. *Über das Conjugations-System der Sanskritsprache in Vergleichung mit jenem der griechischen, lateinischen, persischen und germanischen Sprache* [Sur le système de conjugaison de la langue sanskrite, comparé à celui des langues grecque, latine, persane et germanique]. Francfort
- BOPP, Franz. 1836. *Vocalismus oder Sprachvergleichende Kritiken über J. Grimm's deutsche Grammatik und Graff's althochdeutschen Sprachschatz mit Begründung einer neuen Theorie des Ablauts* [Vocalisme, ou critiques comparatives sur la Grammaire allemande de J. Grimm et le vocabulaire du vieux-haut-allemand de Graff, avec la justification d'une nouvelle théorie de l'apophonie]. Berlin
- DELBRÜCK, Berthold. 1884. *Einleitung in das Sprachstudium – ein Beitrag zur Geschichte und Methodik der vergleichenden Sprachforschung* [Introduction à l'étude du langage – une contribution à l'histoire et méthodologie de la recherche linguistique comparée], 2<sup>e</sup> éd. Leipzig :
- DIEZ, Friedrich. 1836–1843. *Grammatik der romanischen Sprachen* (3 vol.) [Grammaire des langues romanes]. Bonn
- DIEZ, Friedrich. 1853. *Etymologisches Wörterbuch der romanischen Sprachen*. 2 vol. [Dictionnaire étymologique des langues romanes]. Bonn
- FRANÇOIS, Jacques. 2017. *Le siècle d'or de la linguistique en Allemagne – de Humboldt à Meyer-Lübke*. Limoges : Lambert-Lucas.
- GABELENTZ, Georg von der. 1881. *Die Sprachwissenschaft, ihre Aufgaben, Methoden und bisherigen Ergebnisse* [La linguistique, ses tâches, méthodes et acquis actuels]. Leipzig
- HALLIDAY, M.A.K. & Ch. MATTHIessen. 2004. *An Introduction to Functional Grammar* (3th ed.). Londres : Hodder Education.
- HÜLTENSCHMIDT, Erika. 2000. « La professionnalisation de la recherche allemande ». In *Histoire des idées linguistiques* dirigé par S. Auroux. vol.2, 79–96. Bruxelles : Mardaga.
- HUMBOLDT, Wilhelm von. 1823. *Über das Verbum in den Amerikanischen Sprachen* [Sur le verbe dans les langues américaines]. Académie de Berlin. Édité par J. Trabant (1992) : W. von Humboldt, *Über die Sprache*. 82–97. Tübingen & Bâle : Francke-UTB.
- MEYER-LÜBKE, Wilhelm. 1890–1902. *Grammatik der romanischen Sprachen* [4 vol.]. Leipzig : [rééd. Darmstadt 1972, trad.fr. 1895 : Grammaire des langues romanes, par A. et G. Doutrepont, Paris].

- MEYER-LÜBKE, Wilhelm. 1909. *Einführung in das Studium der romanischen Sprachwissenschaft* [Introduction à l'étude de la linguistique romane]. Heidelberg
- MEYER-LÜBKE, Wilhelm. 1911. *Romanisches Etymologisches Wörterbuch* [Dictionnaire étymologique roman]. Heidelberg
- PAUL, Hermann. 1880. *Principien der Sprachgeschichte* [Principes de la linguistique historique]. Leipzig
- SCHLEICHER, August. 1859. *Die deutsche Sprache* [La langue allemande]. Weimar :
- STOROST, Jürgen. 2001. « Die 'neuen Philologien', ihre Institutionen und Periodica : eine Übersicht ». In *History of The Language Sciences / Geschichte der Sprachwissenschaften*, edited by Auroux *et al.* (eds.), 1240–172. Berlin : De Gruyter.
- TESNIERE, Lucien. 1959. *Éléments de syntaxe structurale*. Paris : Klincksieck.