

Patrizia Giampieri
Claudia Mary Forshaw Labruzzo

Technical Translations: a corpus approach for Italian and English speakers

Celid

ARCHITETTURA
ARTI VISIVE
BENI CULTURALI
ECONOMIA
INGEGNERIA
LINGUE
PSICOLOGIA
SCIENZE UMANE



The collection of electronic texts which can be consulted to translate the documents proposed in each chapter is freely downloadable at

www.celid.it/scheda-libro?aaref=1456

Patrizia Giampieri

Adjunct professor of English at the University of Camerino (MC, Italy) and at the Carlo Bo Higher Institute for Interpreters and Translators (Rome). She is currently a PhD student at the Department of Translation, Terminology and Interpreting Studies of the University of Malta. She is a Cambridge Speaking Examiner and a Member of the European Society for Translation Studies. She has authored books on English as a second language, corpus-based translation and English for Special Purposes.

Claudia Mary Forshaw Labruzzo

English/Italian bilingual teacher specialised in Cambridge, IELTS, OET and language exams. She completed an MA in Specialised Translation in 2018 – scientific and institutional texts from Italian and Spanish into English. Since 2014, she has become an English Language Assessment Professional for IELTS and currently collaborates with the NHS, the British National Health System, to help doctors, nurses and physiotherapists prepare for the OET, the English language test for health professionals.

Patrizia Giampieri
Claudia Mary Forshaw Labruzzo

Technical Translations: a corpus approach for Italian and English speakers

The logo for Celid, featuring the word "Celid" in a bold, sans-serif font. A small, stylized icon of a book or document is positioned above the letter "i".

ISBN 978-88-6789-222-8

© 2021 Celid

www.celid.it

celid@lexis.srl

LEXIS Compagnia Editoriale in Torino srl
via Carlo Alberto 55
I-10123 Torino

TABLE OF CONTENT

PREFACE <i>Prof. Maristella Gatto</i>	9
FOREWORD	11
PART I:	
INTRODUCTION TO CORPUS LINGUISTICS, CORPORA AND CONCORDANCERS	
CHAPTER 1 INTRODUCTION AND LITERATURE OVERVIEW	15
1.1. What is computational linguistics?	15
1.2. What is a corpus?	15
1.3. Online and offline corpora	15
1.3.1. Corpus consultation	16
1.4. Collocations and colligations	17
1.5. Lemma and lemmatisation	17
1.6. POS tagging	17
1.7. KWIC	18
1.8. The web as corpus	18
1.9. Web concordancers	18
Bibliography	19
CHAPTER 2 THE WEB AS CORPUS IN PRACTICE	23
2.1. Google	23
2.1.1. Advanced search syntax	23
2.1.2. Collocations	25
2.2. WebCorp	26
2.2.1. Concordances	27
2.2.2. Collocations	28
2.2.3. Lemmatisation	29
2.2.4. Bibliography	30
2.3. The Leeds	30
2.3.1. Concordances	31
2.3.2. The OR operator	31
2.3.3. Lemmatisation	32
2.3.4. POS Tagging	32
2.3.5. Gap between words	34
2.3.6. Collocations	35
2.3.7. The Leeds collection of English corpora	37
2.3.8. Bibliography	38
CHAPTER 3 ONLINE CORPORA IN PRACTICE	39
3.1. The COCA and The BNC	39
3.1.1. Concordances	39
3.1.2. Lemmatisation	40
3.1.3. Collocations	44
3.1.4. POS function	48

3.1.5. The KWIC function	49
3.1.6. Bibliography	50
3.2. The Sketch Engine	50
3.2.1. Concordances	51
3.2.2. Lemmatisation	54
3.2.3. Word's grammar function	55
3.2.4. Collocations and colligations	56
3.2.5. Bibliography	60
CHAPTER 4 OFFLINE CORPUS COMPILATION	61
4.1. Manual compilation	61
4.1.1. Saving doc and pdf files	61
4.1.2. Saving html files	61
4.1.3. Bibliography	63
4.2. BootCaT	63
4.2.1. BootCaT automatic compilation	63
4.2.2. BootCaT semi-automatic compilation	72
4.2.3. Bibliography	76
4.3. AntCorGen	77
4.3.1. Bibliography	79
CHAPTER 5 OFFLINE CORPUS CONSULTATION: CONCORDANCERS	81
5.1. AntConc	81
5.1.1. Uploading the corpus and opening files	81
5.1.2. Generating concordances	82
5.1.3. KWIC function	83
5.1.4. Case-sensitive search	84
5.1.5. Collocations	85
5.1.6. The wildcard character	88
5.1.7. Word list	90
5.1.8. Saving data	91
5.1.9. Bibliography	91
5.2. TextSTAT	92
5.2.1. Creating or opening a corpus	92
5.2.2. Adding files	93
5.2.3. Generating concordances	99
5.2.4. Source file	100
5.2.5. Case-insensitive search	100
5.2.6. KWIC function	101
5.2.7. Collocations	101
5.2.8. The wildcard character	102
5.2.9. Word list	105
5.2.10. Saving data	105
5.2.11. Bibliography	105

**PART II:
CORPUS-BASED TECHNICAL TRANSLATIONS**

FOREWORD	109
CHAPTER 6 BUSINESS	111
Foreword	111
6.1. Business correspondence (offline corpus)	111
6.1.1. Text to translate and corpus compilation	111
6.1.2. Corpus analysis and translation candidates analysis	112
6.1.3. Proposed translation	114
6.1.4. Other ways to carry out the translation task	115
6.2. Business correspondence (online corpus)	115
6.2.1. The Someya non-annotated corpus version	115
6.2.2. The Someya annotated corpus version	117
6.2.3. Other online corpora on business matters	118
6.3. The web as corpus for catalogue products	118
6.3.1. List of products in Italian and English	118
6.3.2. Google advanced search and proposed translations	119
6.3.3. Other domains	123
CHAPTER 7 ECONOMICS AND FINANCE	124
Foreword	124
7.1. The Stock Exchange	124
7.1.1. Text to translate and corpus compilation	124
7.1.2. Corpus analysis and translation candidates analysis	125
7.1.3. Proposed translation	131
7.1.4. Other ways to carry out the translation task	133
7.2. The Hong Kong Financial Services Corpus	133
7.2.1. Hong Kong Financial Services Corpus (non-annotated version)	133
7.2.2. Hong Kong Financial Services Corpus (annotated version)	134
7.2.3. Keys to practice exercises	135
7.3. An Investment Plan for Europe	136
7.3.1. Text to translate	136
7.3.2. Corpus analysis and translation candidates analysis	136
7.3.3. Proposed translation	139
7.3.4. Other ways to carry out the translation task	140
CHAPTER 8 JOURNALISM	141
Foreword	141
8.1. The Grenfell Tower	141
8.1.1. Text to translate and WebCorp settings	141
8.1.2. WebCorp analysis and translation candidates	142
8.1.3. Proposed translation	144
8.1.4. Other ways to carry out the translation task	144
8.2. The Olympic Games	145
8.2.1. Text to translate and WebCorp settings	145
8.2.2. WebCorp, COCA analysis and translation candidates	146
8.2.3. Proposed translation	148

8.2.4. Other ways to carry out the translation task	149
8.3. Climate Change	149
8.3.1. Text to translate and corpus compilation	149
8.3.2. Corpus analysis and translation candidates analysis	150
8.3.3. Proposed translation	155
8.3.4. Other ways to carry out the translation task	156
 CHAPTER 9 MEDICINE	 157
Foreword	157
9.1. Clinical tests	157
9.1.1. Text to translate and corpus compilation	157
9.1.2. Corpus analysis and translation candidates analysis	158
9.1.3. Proposed translation	162
9.1.4. Other ways to carry out the translation task	162
9.2. The SARS-CoV-2 virus	163
9.2.1. Text to translate and corpus compilation	163
9.2.2. Corpus analysis and translation candidates analysis	163
9.2.3. Proposed translation	167
9.2.4. Other ways to carry out the translation task	167
9.3. The Parkinson's Disease	168
9.3.1. Text to translate and corpus compilation	168
9.3.2. Corpus analysis and translation candidates analysis	170
9.3.3. Proposed translation	172
9.3.4. Other ways to carry out the translation task	173

PREFACE

As the title of this book suggests, its main focus is on providing English and Italian teachers, scholars, lecturers, translators and even self-teaching students with a guide to implementing the corpus linguistics approach in the context of technical translation. And it is perhaps no exaggeration, when seeing things from the perspective of translation pedagogy, to maintain that the relatively slow pace at which corpus linguistics has entered the everyday practice of translation training at our Universities over the past two decades is also due to want of books like this. Indeed, there seems to be a greater and greater abundance of research articles that advocate the importance and usefulness of implementing a corpus approach in technical translation training, but - if we exclude some notable exceptions - there is still a relative paucity of books that can practically support trainers and trainees in this challenging itinerary which entails choosing relevant teaching material and appropriate tasks, obtaining access to resources and data, and all this under obvious time constraints on both parts.

Intended for a variety of readers this book thus offers a solution for anyone interested in following a route from theory to practice, making sure that the practical side can be developed to the full. The reader is indeed introduced to all the necessary theoretical background, in chapters that fully address all key issues in corpus linguistics, including those relating to uses of the web as a corpus. The book then guides its readers in all subsequent steps, chapter after chapter, to experiment first-hand with the uses of corpora in technical translation, with the only potential risk of remaining spoilt for choice by plenty of material for use in and outside of the technical translation classroom. Which is obviously a very good risk to run!

Corpus linguistics, corpus-based translation studies, and translation-driven corpora have come a long way over the past few years, and might certainly be the default choice in the future of translation training. Books like this are of great help to make this process smoother and – why not – more pleasant.

Maristella Gatto

FOREWORD

This book explains and shows how corpora and the web as corpus can be consulted successfully to deliver technical translations in a first and second language. It follows a step-by-step approach and provides insights into term search, collocation retrieval and uses of words in contexts of various technical fields. You will become acquainted with online and offline corpora, the web as corpus and will be able to deliver native-like translations in different sectors. The fields tackled by this book are various and include economics, business, medicine and journalism.

This book is intended for teachers, scholars, lecturers, translators and self-teaching students and it has a very practical approach. The first part shows and explains how corpora and the web as corpus can be consulted to deliver technical translations. The first chapter focuses on a literature overview on corpus linguistics. The second chapter explains how to investigate the web as corpus in order to retrieve linguistically reliable information and documents. The third chapter explores how to consult free online corpora, how to use POS (Part Of Speech) tagging and generate word lists, how to write the search syntax accurately and retrieve collocations, etc. The fourth chapter and the final chapter of the first part focus on offline concordancers, which are used to consult offline corpora. The second part of this book follows a practical approach and presents case studies on technical translations tackled by consulting the web as corpus, online and offline corpora. Each chapter of the second part revolves around a specific domain, and shows and explains how to address translations in several technical fields. In particular, it will show how to use free online corpus interfaces, the web as corpus and *ad hoc* offline corpora which are provided with this book.

This book is unique because not only does it clearly show how corpora and the web can be consulted to obtain relevant and reliable information in the field of technical translations, but it also **provides the reader with a collection of electronic texts** which can be consulted to translate the documents proposed in each chapter. In this way, the user will learn how to use corpora successfully by self-practising and will have free databases in economics, business and medicine for his/her own use. For these reasons, this is both a reference book for lecturers or technical translators and a self-study practice guide for students in translation studies. The ideal level of English is at least B2.

**PART I:
INTRODUCTION TO CORPUS LINGUISTICS,
CORPORA AND CONCORDANCERS**

The first part of the book presents corpora and the web as corpus; you will understand how corpora and the web as corpus can be consulted to address technical translations. The first chapter of the first part focuses on a literary overview on computational and corpus linguistics. The second chapter explains how to investigate the web as corpus in order to retrieve linguistically reliable information or documents from the web. The third chapter shows how to consult free online corpora, how to write the search syntax and generate collocations, how to use POS tagging and obtain word lists, etc. The fourth chapter focuses on corpus compilation and the fifth chapter deals with offline concordancers, which are programmes used to consult offline corpora. If you are not acquainted with corpora, corpus compilation and consultation, concordances and/or POS tagging, it is recommendable that you read the first part of the book thoroughly.

CHAPTER 1 INTRODUCTION AND LITERATURE OVERVIEW

In this chapter, you will review the literature on computational linguistics; in particular, you will understand the usefulness of corpora for translation purposes and for second language learning.

Therefore, you will become acquainted with corpora, concordances, node words, collocations, colligations, the web as corpus and many other features.

1.1. What is computational linguistics?

The Association of Computational Linguistics¹ defines computational linguistics as a discipline focused on written and spoken language from a computational perspective. In practice, it blends linguistics with computer science (Schütze & Manning 1999; Jurafsky & Martin 2008). The fields tackled by computational linguistics are broad and include: text summarisation, text retrieval, machine translation, (complex) question answering, use of corpora for linguistics purposes, and many others.

In this book, we will focus on the latter aspect of computational linguistics: corpus composition and consultation.

1.2. What is a corpus?

A corpus (plural: corpora) is a collection of texts (Sinclair 1991) dealing with a particular topic. Nowadays, collections of texts are mostly in electronic format. They are generally gathered for linguistic purposes, such as the analysis of the use of language over the years (Sinclair 1991). Scholars argue that corpora are not merely aimed at collecting texts, but they also represent a language or some part of it (Biber et al. 1998: 246). In particular, they can show language change and evolution over time (Renouf 2002). What is also peculiar and useful about corpora, is that they contain instances of authentic language usage (Gavioli & Aston 2001: 240; McEnery & Xiao 2010: 367; Boulton & Tynes 2014: 301). For these reasons, they are claimed to be particularly useful in language learning (see Giampieri et al. 2019), and in translation training and practice (Biel 2010; Pontrandolfo 2012 and 2017; Vigier-Moreno 2016; Biel et al. 2019; Giampieri 2019b and 2020). As far as translation training and practice are concerned, corpora help deliver accurate translation work and expedite the translation process (Monzó-Nebot 2008; Giampieri 2020). This, however, takes place as long as users are trained in document and information retrieval, and become acquainted with corpus search syntax (Laursen & Pellón 2012, Raído 2014; Zanettin et al. 2014: 11; Giampieri 2019a: 20-21 and 2020).

1.3. Online and offline corpora

Corpora can be online or offline (Anderson & Corbett 2017). Online corpora are ready-to-use collections which are explored by using an online interface. Offline corpora are composed directly by the user and consulted by means of *ad hoc* software (Zanettin 2012; Gatto 2014; Anderson & Corbett 2017: 17). Offline corpora are also referred to as “DIY” (Do It Yourself) (Zanettin 2002), “disposable” (Varantola 2002), “quick-

¹ <https://www.aclweb.org/portal/>

and-dirty” (Tribble 1997), and “*ad hoc*” (Aston 1999; Fletcher 2004). They are built for specific translation projects or for particular aims and can be discarded afterwards. Offline corpora can reach a “saturation point” (Jensen, Mousten & Laursen 2012: 24) beyond which any additional document or text does not add significant information. In this respect, some scholars argue that offline corpora composed of 10 texts per genre yield to reliable results (Williams 1999: 516). This implies that the documents composing offline corpora may not be many, but must be selected carefully. The question of representativeness is, in fact, highly disputed (Tognini-Bonelli 2001: 57), as corpus data will represent specific sector language.

Online corpora can be of different types. For example, there can be parallel, comparable, diachronic and learner corpora (Biel 2010; Pontrandolfo 2012). Parallel corpora are generally composed of source texts and their related translations. Comparable corpora, instead, are composed of original texts in different languages dealing with the same topics (Zanettin 2012: 11; Gatto 2014: 16). Prieto Ramos (2020) highlights the frequent use of parallel corpora for terminological decision making. Many scholars currently report that the differences between parallel and comparable corpora are no longer so clear-cut, especially in international organisations (Fantinuoli and Zanettin 2015: 3-4). Diachronic corpora contain texts of different periods and are used to analyse developments or change in language. Learner corpora are composed of texts produced by the learners of a specific language.

1.3.1. *Corpus consultation*

As anticipated, a corpus is used by means of a specific programme. In online corpora this is generally a user-friendly online interface (see the Sketch Engine platform, Kilgarriff et al. 2014, for example) which users can access for data to be viewed. Simple or complex search queries are, hence, run online. In offline corpora, users resort to a concordancer, such as AntConc (Anthony 2019) or TextSTAT (Hüning 2014), which is a software programme allowing to consult a pre-compiled offline corpus. Generally, the documents composing the corpus are in .txt format.

The system or software which consult either online or offline corpora generates concordance lines (or concordances, Zanettin 2012: 143; Gatto 2014: 23) which are lines, or strings of text showing the searched word placed in the middle of the text. The search word is called a “node word” (Zanettin 2012; Gatto 2014) which is generally highlighted or written in bold to allow for better linguistic analyses.

Corpora have been largely advocated by scholars (Bowker 1998; Zanettin 2012; Granger 2013; Gatto 2014; Anderson & Corbett 2017: 8), as they complement dictionaries (Giampieri 2019a) and tackle for their shortcomings (Tuebert 2002). Many researchers have, in fact, highlighted that conventional lexicographic tools may be insufficient as far as language research and technical translations are concerned (Dominic 2012: 1; Giampieri 2019a and 2020). Corpora, instead, are helpful (Andrades Moreno 2013: 2-3) because they show patterns of language attested usage (Gatto 2014: 29-30), frequent collocations and word uses in context (Jensen, Mousten & Laursen 2012: 30; Vigier Moreno 2016: 107; De Sutter et al. 2017: 27). In this respect, offline corpora are argued to be very effective in specialised translations (Gallego-Hernández 2015: 385), because they focus on specific sector language.

Corpus consultation is reported to increase the translation quality (Pastor & Alcina 2009: 13; Gallego-Hernández 2015: 377; Giampieri 2020: 15) and the user’s self-confidence

(Vigier Moreno 2016: 104; Giampieri 2020: 15). As a matter of fact, corpora help translators make more informed decisions (Rodríguez-Inés 2008, quoted in Vigier Moreno 2016: 104; Vigier Moreno 2019), especially when translating into a second language (see the case studies by Zanettin 2001, by Vigier Moreno 2016 and by Giampieri 2019a and 2020). Corpus consultation, in fact, allows to notice collocations, colligations and word uses in context (Jensen, Moustén & Laursen 2012: 30; Vigier Moreno 2016: 107; De Sutter et al. 2017: 27; Giampieri 2019b: 38).

1.4. Collocations and colligations

As anticipated above, corpora are very precious in language learning and in technical translations because, amongst others, they display collocational use. Firth (1957) described collocations as the tendency of words to keep each other's company (Firth 1957: 11). In practice, collocations show patterns of language usage (Gatto 2014: 29-30) and refer to the frequent co-occurrence of lexical items (Clear 1993, 277; Lehecka 2015). For example, people may say that butter goes rancid, whereas bread goes stale when they "go bad" (Palmer 1981: 77; Baker 2011: 52).

Also, colligations are provided for in corpora. They refer to the co-occurrence of a word with a grammatical class of words (McEnery, Xiao & Tono, 2006: 345). Therefore, differently from collocations, colligations highlight, for example, the relation of a verb with a class of prepositions or adverbs. Colligations also encompass noun pre- or post-modifications (Gatto 2008: 61-64; Gatto 2014: 96-98). For example, investigating whether "the book cover" is more frequent than "the cover of the book" is a matter of colligational use. More precisely, in the first case, the noun "cover" is pre-modified by the noun "book"; whereas in the second case, the noun "cover" is post-modified by the prepositional phrase "of the book". These findings could be very useful to non-native speakers, especially when addressing technical translations (Gatto 2008: 61-64; Gatto 2014: 96-98; Giampieri 2019b).

1.5. Lemma and lemmatisation

A lemma is a headword which includes all the word inflected forms (Anderson & Corbett 2017: 72). Therefore, if the verb "say" is considered as a lemma, its inflected forms "saying", "says" and "said" will be displayed in the search results together with "say".

Not all corpora are equipped with lemmatisation, but if they are, it is possible to retrieve more results and carry out refined research.

1.6. POS tagging

POS tagging, or part of speech tagging, is a process by which every alphanumeric character of a corpus is assigned a tag, or a label. The tag is given on the basis of the word class the alphanumeric character belongs to (Gatto 2014: 17; Anderson & Corbett 2017: 8). For example, if all verbs can be labeled with a V.* tag, all nouns with an N.*, and all proper nouns with an NP tag, it is self-explanatory that it will be very practical to instruct the system to list the nouns following a given verb, or the verbs preceding a specific adverb, etc. Therefore, equipping a corpus with POS annotation is extremely helpful when searching for colligations, collocations, or for specific linguistic formulae.

1.7. KWIC

Online and off-line corpora also show key words in context (KWIC). When corpora have this function, the node word is surrounded by words which are marked or highlighted with different colours (Zanettin 2012: 124). The colours are given on the basis of the word grammatical class. For example, all adjectives in the proximity of the node word may be red, all verbs green, all adverbs blue etc. Alternatively, words may be given a colour depending on their distance from the node word. Therefore, all words in first left position (L1) are blue, those in second left position (L2) are red, etc.

1.8. The web as corpus

An increasing number of Internet users nowadays access the web for language purposes (Zanettin 2012; Gatto 2014). In this case, commercial search engines are used to consult the web considered as corpus. However, users should know how to cope with its “anarchy” (Kilgarriff 2001; Gatto 2014), because they are increasingly confronted with its “gargantuan and uncontrollable” vastness (Gatto 2008: 21). Therefore, search techniques and ways to fine-tune language search must be applied. However, finding adequate and relevant information on the Internet may not be easy, especially for inexperienced users. These can be subject to “cognitive overload” (Sánchez-Gijón 2009: 114; also referred to in Giampieri 2020), which occurs when Internet surfers are not able to put an end to the documentation process (Sánchez-Gijón 2009). This happens because information on the web is now too vast and exceeds users’ needs (Gatto 2008: 32). Furthermore, websites are often unreliable and, therefore, linguistically irrelevant as they are badly written or translated (Cappelli 2007: 98), or they tend to be biased on commercial grounds.

For these reasons, there are several examples in the literature which show how the web can be safely consulted (or crawled) for linguistic research (Krüger 2012; Zanettin 2012; Gatto 2014; Giampieri 2018b; Giampieri et al. 2019). For example, some scholars highlight the importance of using Google advanced search techniques, such as the Boolean operators represented by the “/”, “|” and “-” symbols (Gatto 2008: 54-55; Giampieri et al. 2019: 67). Furthermore, inverted commas can be used to narrow down the search to exact words or phrases (Gatto 2014: 87; Giampieri et al. 2019: 71ff) and the wildcard character “*” represents (or substitutes) any character or word (Zanettin 2012: 124; Gatto 2014: 91). Other advanced search commands concern domain restrictions or exclusions by using the *site:.domain* and *-site:.domain* instructions (Krüger 2012: 517; Giampieri et al. 2019: 70) and the possibility to select the document type to retrieve (e.g., *filetype:pdf*). These commands help narrow down the search and retrieve accurate and targeted documentation from the web (Krüger 2012: 517-518; Zanettin 2012: 124; Giampieri et al. 2019: 69-81).

1.9. Web concordancers

Given the shortcomings of the web as corpus, some scholars developed programmes which crawl the web for linguistic analysis and dispose the results in a way which is suitable for the reader (Gatto 2008: 80; Krüger 2012: 520). These programmes are called web concordances; examples are the WebCorp (Kehoe & Renouf 2002) and the Leeds platforms (Wilson et al. 2010). Web concordancers have user-friendly interfaces which make “raw” data from the web look more intelligible (Zanettin 2012: 66). In this way, they are considered helpful language resources (Gatto 2008: 85) and are used for either translation or L2 teaching purposes (Gatto 2008; Giampieri 2018a, 2018b, 2019a, 2019b,

2019c; Giampieri et al. 2019: 82-100). One of the main drawbacks of web concordancers, however, is the fact that they are grounded in web data. Hence, their results may not be reliable or stable over time (Gatto 2008: 40-41). Furthermore, in some web concordance platforms it is not possible to restrict the search to specific domains, which makes the results untargeted (Giampieri 2018b: 47). Other web concordancers, instead, are not equipped with POS tagging, which limits the possibility to narrow down the search and to find patterns of language attested usage (Giampieri 2018a: 45; Giampieri 2019c: 119). Therefore, it is evident that web concordancers are useful tools for language learning and translation purposes. However, they should be complemented by and used in conjunction with other language resources (Giampieri 2018b: 51; Giampieri 2019c: 119).

Bibliography

- Anderson, W. and Corbett, J. 2017. *Exploring English with Online Corpora*. Second Edition. Basingstoke: Palgrave Macmillan.
- Andrades Moreno, A. 2013. Internet como fuente para la compilación de corpus jurídicos. *Enlaces* 15: 1–15.
- Anthony, L. 2019. *AntConc, build 3.5.8*. Center for English Language Education in Science and Engineering, School of Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan Help file version: 001. <https://www.laurenceanthony.net/software/antcon/>
- Aston, G. 1999. Corpus use and learning to translate. *Textus* 12 (2): 289–314.
- Baker, M. 2011. *In other words: a course on translation*. Second Edition. London/New York: Routledge.
- Biber, D., Conrad, S. and Reppen, R. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biel, Ł. 2010. Corpus-Based Studies of Legal Language for Translation Purposes: Methodological and Practical Potential. In C. Heine and J. Engberg (Eds.), *Reconceptualizing LSP. Online proceedings of the XVII European LSP Symposium 2009*, 1-19.
- Biel, Ł., Engberg, J., Martín Ruano, R. and Sosoni, V. 2019. *Research Methods in Legal Translation and Interpreting: Crossing Methodological Boundaries*. London: Routledge.
- Boulton, A. and Tyne, H. 2014. Corpus-based study of language and teacher education. In M. Bigelow and J. Enns-Kananen (Eds.), *The Routledge Handbook of Educational Linguistics*, 301-312. New York: Routledge.
- Bowker, L. 1998. Using specialized monolingual native-language corpora as a translation resource: A pilot study. *Meta* 43 (4): 631–651.
- Cappelli G. 2007. The translation of tourism-related websites and localization: problems and perspectives. In A. Baicchi (Ed.), *Voices on Translation, RILA Rassegna Italiana di Linguistica Applicata*, 39 (1-2): 97-115.
- Clear, J. 1993. From Firth principles: computational tools for the study of collocation. In M. Baker, G. Francis and E. Tognini-Bonelli (Eds.), *Text and Technology*, 271-292. Amsterdam: John Benjamins.
- De Sutter, G., Cappelle, B., De Clercq, O., Loock, R., and Plevoets, K. 2017. Towards a corpus-based, statistical approach to translation quality: Measuring and visualizing linguistic deviance in student translation. *Linguistica Antverpiensia, New Series: Themes in Translation Studies* 16: 25–39.
- Fantinuoli C. and Zanettin F., 2015, Creating and using multilingual corpora in translation studies. In Fantinuoli C. and Zanettin F. (Eds.), *New directions in corpus-based translation studies*, 1-10. Berlin: Language Science Press.
- Firth, J. R. 1957. *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Fletcher, W. H. 2004. Facilitating the compilation and dissemination of ad-hoc web corpora. In G. Aston, S. Bernardini and D. Stewart (Eds.), *Corpora and Language Learners*, 273–300). Amsterdam: John Benjamins.

- Gallego-Hernández, D. 2015. The use of corpora as translation resources: A study based on a survey of Spanish professional translators. *Perspectives* 23 (3): 375-391.
- Gatto, M. 2008. *From body to Web*. Bari: Laterza.
- Gatto, M. 2014. *Web as corpus: Theory and Practice*. London: Bloomsbury.
- Gavioli, L. and Aston, G. 2001. Enriching reality: Language corpora in language pedagogy. *ELT Journal* 55 (3): 238-246.
- Giampieri, P. 2018a. *Inglese giuridico, web e corpora*. Milan: Giuffrè.
- Giampieri, P. 2018b. The web as corpus and online corpora for legal translations. *Comparative Legilinguistics* 33: 35-55.
- Giampieri, P. 2019a. Corpus-based translation in the tourism sector: a case study with final year bachelor students. *MediAzioni* 24: 1-32. <http://mediazioni.sitlec.unibo.it>
- Giampieri, P. 2019b. The BoLC for legal translations: a trial lesson. *Comparative Legilinguistics* 39: 21-46.
- Giampieri, P. 2019c. The web as corpus in ESL classes: A case study. *International Journal of Language Studies* 13 (2): 91-108.
- Giampieri, P., Campa, S. and Milani, G. 2019. *La didattica delle lingue straniere attraverso il web*. Naples: Simone.
- Giampieri, P. 2020. Volcanic experiences: comparing non-corpus-based translations with corpus-based translations in translation training. *Perspectives* 28 (1): 1-18. DOI: 10.1080/0907676X.2019.1705361
- Granger, S. 2013. *Learner English on Computer*. Abingdon: Routledge.
- Hüning, M. 2014. *TextSTAT - Simple Text Analysis Tool*. Freie Universität Berlin. Institut für Deutsche und Niederländische Philologie - Niederländische Philologie / Sprachwissenschaft. <http://neon.niederlandistik.fu-berlin.de/en/textstat/>
- Jensen, V., Moustén, B., and Laursen, A. L. 2012. Electronic Corpora as Translation Tools: A Solution in Practice. *Communication & Language at Work* 1 (1): 21-33.
- Jurafsky, D. and James, M. 2008 *An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, Second Edition. Prentice Hall.
- Kilgarriff, A. 2001. Web as corpus. *Proceedings of the Corpus Linguistics Conference (CL 2001)*. Lancaster University, UK.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. 2014. The Sketch Engine: ten years on. *Lexicography* 1: 7-36.
- Kehoe, A. and Renouf, A. 2002. WebCorp: Applying the Web to linguistics and linguistics to the Web. *Proceedings of the WWW 2002 Conference*. Hawaii: Honolulu, 7-11 May 2002.
- Krüger, R. 2012. Working with corpora in the translation classroom. *Languages in the globalised world: Interpreting multilingual literacy and cross-cultural encounters. Special issue of Studies in Second Language Learning and Teaching* 2 (4): 505-525.
- Laursen, A. L., and Pellón, I. A. 2012. Text Corpora in Translator Training. *The Interpreter and Translator Trainer* 6 (1): 45-70.
- Manning, C. and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*, Cambridge, Massachusetts, USA. MIT Press.
- McEnery, T., Xiao, R. and Tono, Y. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. New York: Routledge.
- McEnery, T. and Xiao, R. 2010. What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning. Vol. 2*, 364-380. London & New York: Routledge.
- Monzó-Nebot, E. 2008. Corpus-based activities in legal translator training. *The interpreter and translator trainer* 22: 221-252.
- Palmer, F. R. 1981. *Semantics*. Cambridge: Cambridge University Press.
- Pastor V. and Alcina A. 2009. Search techniques in corpora for the training of translators. In I. Ilisei, V. Pekar, and S. Bernardini (Eds.) *Proceedings of the International Workshop on Natural Lan-*

- guage Processing Methods and Corpora in Translation, *Lexicography and Language Learning*, 13-20. Bulgaria: Incoma.
- Pontrandolfo, G. 2012. Legal Corpora: an overview. *Rivista Internazionale di Tecnica della Traduzione* 14: 121-136.
- Pontrandolfo, G. 2017. Los corpus como recursos en la enseñanza del español jurídico: aprender descubriendo. In Espaliu Berdud, C., De Miranda Vazquez, C.; Jiménez Yañez, R. (eds.), *¿Cómo la innovación mejora la calidad de la enseñanza del derecho? Propuestas en un mundo global*. Pamplona: Aranzadi, 231-252.
- Prieto Ramos, F. 2020. The use of resources for legal terminological decision-making: patterns and profile variations among institutional translators. *Perspectives*, 29(2), pp. 278-310.
- Raído, V. E. 2014. *Translation and Web Searching*. New York: Routledge.
- Renouf, A. 2002. The Time Dimension in Modern English Corpus Linguistics. In B. Kettemann and G. Marko (Eds.), *Teaching and Learning by Doing Corpus Analysis: Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz 19-24 July 2000*, 27-42. Amsterdam / New York: Rodopi.
- Rodríguez-Inés, P. 2008. Translation into L2 in education and in real life, and how electronic corpora can help. In *Presentation delivered at the Eighth Translation Conference: The Changing Face of Translation*, Portsmouth: University of Portsmouth.
- Sánchez-Gijón, P. 2009. Developing documentation skills to build do-it-yourself corpora in the specialised translation course. In A. Beeby, P. Rodríguez-Inés, and P. Sánchez-Gijón (Eds.), *Corpus use and translating. Corpus use for learning to translate and learning corpus use to translate*, 109-127. Amsterdam: John Benjamins.
- Sinclair, J. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Stewart, D. 2012 *Translating Tourist Texts from Italian to English as a Foreign Language*, Naples: Liguori.
- Teubert, W. 2002. The role of parallel corpora in translation and multilingual lexicography. In B. Atenbergh and S. Granger (Eds.), *Lexis in Contrast*, 189-214. Amsterdam/Philadelphia: John Benjamins.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.
- Tribble, C. 1997. *Improvising corpora for ELT: Quick-and-dirty ways of developing corpora for language teaching*. Paper presented at Practical Applications of Language Corpora, Lodz, Poland.
- Varantola, K. 2002. Disposable corpora as intelligent tools in translation. *Cadernos de Tradução IX – Tradução e Corpora* 1 (9): 171-189.
- Vigier Moreno, F. J. 2016. Teaching the Use of ad hoc Corpora. *Language and Law / Linguagem e Direito* 3 (1): 100-119.
- Vigier Moreno, F. J. 2019. Corpus-assisted translation of specialised texts into the L2. From the classroom to professional practice. *Trans-Kom* 12 (1): 90-106.
- Williams, I. A. 1999. Corpus-based contrastive analysis and the translation of specialized texts. In M. Á. Vega and R. Martín-Gaitero (Eds.), *Lengua y cultura: Estudios en torno a la traducción*, 515-522. Madrid: Universidad Complutense.
- Wilson, J., Hartley, A., Sharoff, S., and Stephenson, P. 2010. Advanced corpus solutions for humanities researchers. In R. Otoguro, K. Ishikawa, H. Umemoto, K. Yoshimoto and Y. Harada (Eds.), *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, 769-778.
- Zanettin, F. 2001. Swimming in Words: Corpora, Translation, and Language Learning. In G. Aston (Ed.), *Learning with corpora*, 177-197. Houston, TX: Athelstan
- Zanettin, F. 2002. Corpora in translation practice. In E. Yste-Rodrigo (Ed.), *Language resources for translation work and research, LREC 2002 Workshop Proceedings, Las Palmas de Gran Canaria*, 10-14.
- Zanettin F. 2012. *Translation-Driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. London: Routledge.
- Zanettin, F., Bernardini, S., and Stewart, D. 2014. *Corpora in translator education*. New York: Routledge.