Corpus-based Translation for Research, Practice and Training (Topics in Translation Series) - Mona Arhire. 2014.
Iaşi: Institutul European, 253 pages.
ISBN 978-606-24-0037-8

Reviewed by Raluca SINU[1]

The corpus has represented a breakthrough in research in all language related fields, and Translation Studies is no exception. It is easy to understand then why so many articles, books, conferences in recent decades have dealt with this topic. The volume *Corpus-based Translation for Research, Practice and Training* is part of this trend and discusses the role of corpus in the field of Translation Studies. As its author argues, the book "is conceived so as to foster the translation professionals' embracing the modern and promising approach to Translation Studies based on corpora" (back cover). It addresses aspects pertaining both to the theory and the practice in the field, in an attempt to prove that "practitioners and trainers in the field of translation can benefit from engaging with corpora" (p. 15).

Before we discuss the structure of this book, it is worth noting that it is built around the concept of corpus, each of its chapters providing answers to a corpus-related question: what is the impact of corpus on translation studies? (Chapter One and Chapter Two), what is a corpus? (Chapter Three), what are the criteria for compiling or assessing a corpus? (Chapter Four and Chapter Five) what are the stages of corpus compilation? (Chapter Seven) what are some of the applications of the corpus for translation researchers, practitioners and trainers? (Chapter Six, Chapter Eight and Chapter Nine) The following paragraphs will offer an overview of the individual chapters.

In the "Introduction", the author opens the discussion about the field of Corpus-based Translation Studies (CBTS). She highlights the emergence and development of CBTS, a field which stands at the crossroads of translation studies, corpus linguistics, and computational linguistics, while pointing out that this has not been an easy partnership considering the sometimes poor relationship between corpus linguistics and translation studies, the latter being rarely the focus of research conducted by the former. Arhire also links the development of CBTS to the advent of the electronic corpus, which revolutionized translation studies. Through the electronic tools for corpus analysis, corpus-based research encouraged "the systematic investigation of the translational language and patterning, on sizeable

---

[1] Transilvania University of Brasov, raluca.sinu@unitbv.ro

corpora, leading to more reliable results" (p. 24). The chapter also stresses the gap between the Romanian and the international contexts in terms of the development of CBTS, calling for more Romanian fundamental research in the area of corpus-based methodology.

In Chapter Two, "The contribution of Corpus-based Translation Studies to the interdisciplinary character of Translation Studies", it is argued that CBTS strengthened the interdisciplinarity of Translation Studies, because it placed this field in close contact with corpus linguistics and information technology, since CBTS "investigates translational data with methodological practices typical of Corpus Linguistics joined with software tools" (p. 32). The author further exploits the role of the corpus as a mediator between Translation Studies and Corpus Linguistics, on the one hand, and Translation Studies and Information Technology, on the other. The corpus and language analysis tools borrowed from Corpus Linguistics supported descriptive translation studies, whereas the partnership between translation studies and computer sciences opens the way for unprecedented possibilities, which extend to the practical side.

Chapter Three, "Corpus: definition, use, typology", focuses on the key concept of the book, i.e. the corpus. In CBTS, the corpus is defined as "a collection of texts held in machine-readable form and the same format, and capable of being analyzed automatically and semi-automatically" (p. 42). Arhire goes on to present the types of corpora specific to CBTS, namely the *parallel corpus*, the *comparable corpus*, and the *learner corpus*. The first one is a collection of texts and their translations, while the second includes similar texts in two or several languages, or a sub-corpus of texts produced by native speakers and one made up of texts translated into the same language. The learner corpus can be either parallel or comparable, and what distinguishes it from the other types is the fact that it is made up of learner language samples and compiled for educational purposes. Then, the author also introduces other classifications of corpora based on criteria such as medium (printed vs. electronic corpus), number of languages involved (monolingual, bilingual or multilingual corpus), domain (general vs. specialized corpus), size (closed vs. open-ended corpus), temporality (synchronic vs. diachronic corpus), etc., describing them and providing examples. The classification criteria applied are not mutually exclusive, this is why the categories obtained sometimes overlap, for example parallel or comparable corpora can be, at the same time, synchronic or diachronic, general or specialized, etc.

Chapter Four and Chapter Five are constructed by the author as complementary, both of them dealing with issues pertaining to corpus design, the first stage in corpus research. Chapter Four, "Corpus design: the compiler's perspective", presents the factors at play in corpus collection. The author invites the reader to put on the compiler's hat and decide how to build the corpus, in other words what texts the corpus should include and what criteria should be taken into consideration. The author underlines a number of features a corpus should have in

order to be distinguished from a random collection of texts. Thus, the corpus should be representative for the type of text under investigation, balanced, large enough, and reliable. Other decisions refer to the time span the corpus should cover, and to whether full texts or samples should be included. In addition to these corpus-internal factors, Arhire also discusses external constraints and risks concerning timeliness, ethical issues, the text sources, the funding, even the status of translation studies, which can shape the corpus.

The point of view changes in the next chapter, "Corpus design: the user's perspective", from the process of corpus compilation to that of corpus assessment. The users of a ready-made corpus, either researchers or translators, should base their evaluation on features such as the purpose and audience of corpora, their reliability, accessibility, structure and content. Noteworthy is the overview of corpora including the Romanian language (§5.3.1), which supplies interesting examples of corpus projects and their compilation, for example RACAI, ROCO, RoWaC, ROMBAC, ARS-ROCOCO, etc., although few of them are actually designed for translation research purposes. This is why the author believes the Romanian language is still underrepresented in CBTS, which prompts her to call for more initiatives in this field.

> [I]t is important that both Romanian theorists and practitioners feel encouraged to commit themselves to continuing or taking up corpus-based projects. The area of Corpus-based Translation Studies is still underdeveloped. The cooperation at the academic level in this interdisciplinary study area should bring together teams of linguists and of specialists in information technology in order to achieve pertinent, high quality research. (p. 105)

At first sight, Chapter Six, "Translation Universals", seems unrelated to the previous ones. However, a closer look reveals the connection: translation universals represent one of the major results of corpus-based translation research, even before the electronic age and the emergence of CBTS. Despite the lingering controversy surrounding the idea of universals in translation, many authors believe, without any doubt, in the existence of "a series of features which the target texts exhibit irrespective of the languages involved" (p. 112). Out of the list of translation universals available in the specialized literature, the author discusses in more detail explicitation, simplification and normalization, quoting many studies in this field. Some of them are based on corpora, for example, CTF (Corpus of Translated Finnish) (p. 120), a corpus of contemporary German texts and their translations into English (p. 122), but most are based on smaller scale empirical research.

Chapter Seven, "Research methodology for Corpus-based Translation Studies", starts by describing the stages involved in research conducted with the help of corpora in general, and CBTS in particular. Thus, the normal workflow

usually includes: deciding on the purpose of the research, compiling or finding a suitable corpus, hypothesizing, preparing the corpora for analysis, and the quantitative and/or qualitative analysis of the material. For the second stage, i.e. creating or identifying a suitable corpus (§7.1.2), information was already provided in chapters four and five, which looked at what makes a good corpus, the criteria compilers and users have to keep in mind when creating/choosing a corpus. The discussion concerning the hypothesizing stage (§7.1.3) revolves around research questions formulated in translation studies, some of them already presented as translation universals in Chapter Six. The last two stages in the process are the most terminologically complex, as the author introduces and explains a series of terms related firstly to formatting the texts in the corpus, e.g. annotation, tagging, lemmatization, markup, and secondly to the analysis of the corpus, e.g. concordancing tools, frequency lists, etc. The examples provided here are very helpful in understanding these concepts borrowed from corpus linguistics and in illustrating the link between theory and practice.

In the second part of Chapter Seven, the author surveys some of the latest tools and well-known topics in corpus-based research. Thus, she examines the case of Wordsmith tools, Corpus Presenter, Tropes, TextSTAT, WordCruncher Viewer, etc., and their usefulness for the analysis of translation universals, as well as individual variations in translation, translation norms and conventions, the relation between translation and language change, interpreting, learner corpora, etc. The author mentions that the research can be extended to all language levels (lexical, morpho-syntactic, stylistic, discoursive, etc.), a feature which makes it very tempting for the researcher. At the end of the chapter, she also provides valuable information concerning the tools employed in the investigation of Romanian corpora or corpora in which Romanian is half of the language pair (§7.4), stressing the need for more involvement from linguists in this area, dominated so far by IT specialists.

If Chapter Seven is aimed at researchers in the field of translation studies, the next two chapters address practitioners and trainers, respectively. Chapter Eight, "Corpora as resources for professional translation", highlights the advantages of using corpora at various stages in the translation process such as documentation, lexical search and acquisition, etc. For example, in documentation, the corpus "can be a reliable and fast source of information, facilitating the translator's acquiring content knowledge and specialized terminology" (p. 172). Also, with the help of the electronic tools associated to it, a corpus can supply information on the meanings of lexical items, as well as on their grammatical behaviour. In addition, among other things, it can be the basis for creating glossaries, for studying an author's style, for raising the translator's awareness to professional translation strategies (§8.1.3).

The last chapter of the book, "Corpus-based translator training", focuses on the (potential) role of corpora in the training of translators. Arhire begins by underlining the importance of both theory and practice in the training of future translation professionals. The theoretical training introduces the future translator to

the literature in the field, to the results of the existing research, thus confirming the interdependence of training, on the one hand, and practice, theory and research, on the other, in translation studies. The corpora-based case studies presented in sections 9.3 and 9.4 are useful for those teaching translation as they highlight specific problems encountered in the context of translation training and also suggest solutions. The author expresses her views on translation training based on her professional experience and her in-depth knowledge of the Romanian context. She advocates for the introduction, on a larger scale, of translator training programmes at institutional level in Romania:

> [I]t is a fact that corpus analysis tools have been widely used for research purposes worldwide. It seems though that in translator training the systematic use of these tools as translation aids has not yet been sufficiently explored. Viewed in this global context, the Romanian translational environment has been relatively little involved in both facets of corpus translation although it is clear that access to corpora, together with handy software tools, represents a means to improve the trainees' language and cultural competences as well as their translation skills. (p. 198)

The book also contains a rich section of appendices which offers a list of corpora, many of them mentioned in the text, and the links to their websites (Appendix 1), a list of monolingual and multilingual corpora with the Romanian language (Appendix 2), a list of some of the most widely used software tools for corpus creation and analysis at the international and national levels (Appendices 3 and 4). The last appendix includes a form for the assessment of online corpora used in a small-scale research conducted by the author at her home university, i.e. *Transilvania* Universty of Braşov, an occasion to have another look at the criteria for corpus assessment presented in Chapter Five. In addition to the author's experience in translation practice and translator training, the book is informed by the latest studies, without neglecting the fundamental works in this field, as reflected by the extensive bibliography.

Perhaps the most important contribution of the book resides in its constant references to the Romanian context. Several chapters bring into discussion the Romanian situation, drawing attention to the lack of corpus-based research (§1.3), presenting the results of a corpora assessment project conducted by the author with the help of MA students from the Faculty of Letters at *Transilvania* Universty of Braşov (§5.1), listing the corpora which include the Romanian language (§5.3.1), as well as the electronic tools available for the corpus-based analysis of Romanian and the projects conducted with these tools (§7.4). At the same time, the author surveys the available literature on this topic, quoting some of the most relevant works such as *Limbaj şi tehnologie* [Language and Technology] (1996), edited by D. Tufiş; *Speech Technology and Human-Computer Dialogue* (2003), edited by C. Burileanu;

Resurse lingvistice şi instrumente de prelucrare a limbii române [Linguistic Resources and Tools for Processing the Romanian Language] (2006), edited by C. Forăscu, D. Tufiş and D. Cristea; or the *Proceedings of the 8th International Conference "Linguistic Resources and Tools for Processing of the Romanian Language*" (2012). Thus, the author offers a broad description of the current state of corpus-based research in Romania.

We believe the book fulfills the goal set by its author, that of delivering an overview of Corpus-based Translation Studies, while helping close a gap in the Translation Studies literature published in Romania. It demonstrates beyond any doubt the potential of corpus-based research and its advantages for translation studies, encouraging its readers to continue research in this area.

References

Burileanu, C. (ed.) 2003. *Speech Technology and Human-Computer Dialogue*. Bucharest: Editura Academiei Române.
Forăscu, C., D. Tufiş and D. Cristea (eds.) 2006. *Resurse lingvistice şi instrumente de prelucrare a limbii române*. Iaşi: "Alexandru Ioan Cuza" University Publishing House.
Moruz, M.A., D. Cristea, D. Tufiş, A. Iftene and H. N. Teodorescu (eds.) 2012. *Proceedings of the 8th International Conference "Linguistic Resources and Tools for Processing of the Romanian Language"*. Iaşi: "Alexandru Ioan Cuza" University Publishing House
Tufiş, D. (ed.) 1996. *Limbaj şi tehnologie*. Bucharest: Editura Academiei Române.