

***Corpus-Based Translation
for Research, Practice and Training,***
Mona Arhire
Institutul European, Iași, 2014, 254 p.

Daniela HĂISAN
“Ștefan cel Mare” University, Suceava, Romania

Originating in the mathematical computability theory, the all-pervading computer metaphor has of late become so commonplace we no longer see in it a figure of speech; after all, as long as wearable technology is perceived as a perfectly acceptable and natural extension of the body, likening the mind to the software and the brain to the hardware of a computer is as natural as breathing. Still, when we get down to brass tacks, softwares always come in handy when the “human software” feels overwhelmed with unmanageable tasks. When it comes to juxtaposing translation and technology, apart from “machine translation” which immediately, virally, springs to mind, there is also an entire list of electronic tools that have already proven their efficiency and the multiple benefits of this quasi-symbiotic relationship.

Corpus-Based Translation for Research, Practice and Training, the book under review, certainly fills a gap (as it constantly keeps reminding the reader), at least as far as Romania is concerned, at the crossroads of corpus research and translation studies. One of the newest Topics in Translation Series, Mona Arhire’s book is adequately and thoroughly equipped: a list of figures and tables, another of abbreviations and acronyms, a series of due acknowledgements, and a foreword provide a solid

vanguard, whereas the six appendices, the bibliography and Web references give the volume both symmetry and balance.

The Introduction, which is also the first chapter, offers an instructive glimpse into the history of CTBS (Corpus-Based Translation Studies). We find out the new interdiscipline is but two decades old and shares a common lineage with DTS (Descriptive Translation Studies). As a matter of fact, for the sake of precision, Mona Baker is identified as, literally, “the mother of CBTS” (p. 22, q. Laviosa, 2002) for having put corpus research forward as methodology in TS.

The second chapter deals with *The Contribution of Corpus-Based Translation Studies to the Interdisciplinary Character of Translation Studies*. After a number of definitions of interdisciplinarity emphasizing both its pros and cons, the author proceeds to divide it into *internal interdisciplinarity* (which refers to the relation of TS with other disciplines pertaining to the wide area of letters, languages and literatures, such as applied linguistics, contrastive linguistics, literary and cultural studies, discourse analysis, stylistics, pragmatics, lexicography, terminology etc.) and *external interdisciplinarity* (which necessarily involves two aspects: the contact of TS with a variety of text-types belonging to the most diverse areas of specialization, on the one hand, and the partnership with the field of computer science, on the other). In doing so, she makes the very essence of CTBS appear more clearly in so complex an area with fuzzy boundaries.

By the end of the chapter we run into: “It all starts with the compilation of texts to be included in a corpus.” (p. 36), which anadiploically paves the way for the next section, *Corpus: Definition, Use, Typology*. Here, the reader gets acquainted with the general characteristics corpora need to possess: sampling and representativeness, finite size, a machine-readable form and a standard reference (p. 42), with a new terminology (corpora used

in TS are generally called “translation-driven corpora”) and an intricate taxonomy based on a variety of criteria. Apart from the basic *parallel* (or *bilingual*) *corpus* (commonly regarded as a collection of source-language texts and their translations), the *comparable corpus* and the *learner corpus*, we get to classify corpora by medium (*printed* or *electronic*), origin (*native* or *non-native* electronic texts), domain or content (*reference / general* versus *specialized / terminological* corpora), size (*closed / static* versus *open-ended / dynamic / monitor corpora*), directionality – between a given source and its target, either one way or round trip (*unidirectional / monodirectional* versus *bidirectional / multidirectional* corpora), temporality (*synchronic* or *diachronic*), mode (*written* or *spoken*), status (*published* or *unpublished*), the translator’s status (*professional* or *trainee*) etc. One of the key features to take into account is the similarity that the sub-corpora making up the corpus needs to display.

Chapters 4 and 5 tackle the issue of corpus design from two different perspectives: the compiler’s and the user’s (the two do not always coincide, we are meant to deduce, except in the case of self-made corpora). What compilers are supposed to bear in mind when making up a corpus is representativeness (which determines the kinds of research questions that can be addressed as well as the generalizability of the results of the research (p. 69), followed by reliability, authenticity and comparability. Last but definitely not least, the question of copyright (in this particular case, tracing who holds copyright for both source texts and translations and obtaining permission to reproduce the material), treated in the subchapter entitled *Ethical Issues*, is not only a must, but a time-consuming stage worth considering from the very beginning. The user’s perspective, on the other hand, relies more on the structure and content of the corpus, and the author of the book sees fit to enlarge mainly upon the *Corpora Representing the Romanian Language* (subchapter 5.3.1.). Although we are

introduced to a number of acronyms: ROMBAC (one of the most important corpora of Romanian created within the Research Institute RACAI), ARS-ROCOCO, ACCURAT, RoGER etc., the author concludes regretfully that the Romanian language is still underrepresented.

The sixth chapter, dedicated to *Translation Universals*, speaks about translation as a “literary genre apart” (Ortega y Gasset), as a “third code” (Frawley) or even a “third language” (Duff). Considered by some as vague and rather indefinite in meaning (Anthony Pym, 2010), the notion of “universals”, put forward and then partially retracted by Mona Baker, is split by Chesterman (2004) into *S-universals* (based on a parallel corpus made up of two sub-corpora, one of source-texts and the other of target-texts) and *T-universals* (grounded on the investigation of a comparable corpus comprising a sub-corpus of translated texts and another one including native language). As for Baker, she first (1993) identifies six features common to all translations, namely explicitation, simplification, normalization, avoidance of repetitions, naturalization and the existence of common features in target language texts originally produced in the target language. Later on (1996), she suggested only four of the universals, namely simplification, explicitation, normalization and leveling out, to come to the realization, a few years later, that the term “universals” is not suitable to refer to typical patterns of translation.

The so-called “universals” are nevertheless described in full detail. Explicitation, for one, the easiest to assess from a quantitative point of view, is further classified by Klaudy (1996) into *obligatory* (originating in the structural differences between languages), *optional* (attributed to differences in text-building strategies and stylistic preferences), *pragmatic* (due to the cultural differences between SL and TL) and *translation-inherent* or *translation-proper* explicitations (which derive from the translation process *per se*) (p. 116) Simplification sometimes occurs in the target

language due to the lack of culturally matching terms in the languages involved in translation; when it is inherently stylistic, it involves breaking up long sentences, omissions of repetitions, shortening of complex sentences and the use of common language. Normalization, which can also be viewed in relation to simplification, has, according to Scott (q. in Laviosa, 2002) two poles: one due to the systemic constraints of the target language and the other resulting from the translator's own preferences. Finally, leveling out is defined by Mona Baker as the tendency to "steer a middle course between any two extremes, converging towards the centre" (1996: 184). The chapter ends with a very interesting, though very brief, note on the so-called *Unique Item Hypothesis* (UIH) (Tirkkonen-Condit, 2002, 2004), which is the tendency of TL items with no equivalents in the SL to be underrepresented in translated texts as compared to comparable originals.

Chapter 7 goes straight to the heart of the problem by identifying and describing the main stages in the research methodology: defining the intention or purpose of the research; compiling or finding the suitable corpus; hypothesizing; preparing the corpus for analysis: format, annotation, tagging; quantitative and qualitative corpus analysis. It is, again, Mona Baker (1993) who provides a list of hypotheses that might be used for the study of the translations comprised in a corpus. One of them, for instance, is the premise that translated texts tend to exhibit excessive use of features that are typical for the target-language style; another, that translations have a tendency to display a higher level of disambiguation simplification; sometimes translated texts tend to be more explicit than their pair source-language texts, or, quite understandably, to omit repetitions.

To be put to good use, corpora also need to be carefully annotated, either procedurally (encoding information about the visual formatting of texts: font, size etc.) or structurally (the linguistic and extra-linguistic content: bibliography, logical

structure, morphemes etc.). Various kinds of tag systems, lemmatizers (*lemmatization* being the process of grouping together words which are related by inflection or derivation), markups, statistical measurement tools (such as the *Wordsmith tools*, able to perform calculations of “keyness” of keywords) and concordancers (which enable the study of collocations and colligations) greatly contribute to increasing the efficiency of the research. Again, although mention is made of Romanian corpora and tools (for example, a small-size parallel corpus of general use English texts and their translations into Romanian used to investigate the formation of *neoterms* from a translational perspective or a multilingual parallel corpus made up of George Orwell’s novel *1984* and its translations into several languages, Romanian included), statistics lead to the (partial) conclusion that so far, “the research of Romanian has been mostly undertaken by IT specialists, within the branch of computational linguistics with little involvement of linguists” (p. 165).

Electronic corpora will undoubtedly prove useful both as *Resources for Professional Translation* (as shown in chapter 8) and as part of the *Translator’s Training* (chapter 9), whether we are talking about small-range or large-scale empirical research. The book being reviewed is no less profitable, despite its patchwork quality (a side effect of it being essentially a monograph) and its occasional but definitely condonable stylistic awkwardness.

Note: This contribution is a part of the exploratory research programme Traduction culturelle et littérature(s) francophones: histoire, réception et critique des traductions, CNCS PN-II-ID-PCE-2011-3-0812 Contract 133/27.10.2011.