УДК 81.33

Corpus Use in the Translation Classroom, or Dies Diem Docet

Valentina A. Kononova*
Siberian Federal University
79 Svobodny, Krasnoyarsk, 660041, Russia

Received 20.10.2014, received in revised form 12.11.2014, accepted 06.12.2014

This article discusses the advantages of language corpora use in the modern university translation classroom. Corpora can be valuable resources for translation students and also a solid base for development of professional translator competence. The rationale for using a corpus as a valuable teacher/ learner recourse and a learner tool is outlined. The article traces links between work in the corpus linguistics community and the world of practicing translators. A few suggestions are put forward in order to encourage a wider discussion on challenges in translation education.

Keywords: language corpus, corpora, the Russian National Corpus, translation competence, concordancer, corpus-related activities.

Research area: philology.

Introduction

The relationship between teaching translation and language corpora has recently become the topic of numerous publications (Leech 1997, Plungyan 2007, Grabovsky 2007, Shmeleva 2007, Stewart 2009, Frankenberg-Garcia 2009, Sánchez-Gijón 2009, Laviosa 2010, etc.). Corpus studies “boomed from 1980 onwards, as corpora, techniques and new arguments in favour of the use of corpora became more apparent. Currently this boom continues – and both of the ‘schools’ of corpus linguistics are growing” (McEnery & Wilson, 2001: 24).

Corpus-based methodology and research have been successfully brought forward into the state-of-the-art teaching methods and strategies in the past two decades and have undeniably had a considerable impact on language and translation teaching in the world. The reminder of Mossop “if you can’t translate with pencil and paper, then you can’t translate with the latest technology” sounds noteworthy, as well as fully automatic translation programmes remain a chimera (Beeby et al, 2009). However, language corpora have been undoubtedly added to the technology-based range of resources at the translator’s disposal.

Some records on corpora use in Russia and Europe

At the moment, we cannot place that on record talking about Russian universities. Here remains a gap between what applied corpus linguistics can offer and what has been done (or not) with corpora in real teaching practice. This

* Corresponding author E-mail address: v.kononova@mail.ru

phenomenon could be explained and excused by a comparatively recent appearance of the Russian National Corpus – it was created within RAS (Russian Academy of Sciences) programme "Philology and Informatics" by scholars from a number of RAS institutes with the participation of Moscow, St. Petersburg, Voronezh and Izhevsk universities and introduced on Yandex in April 2004 (http://ruscorpora.ru). From there, the Russian language became another language in the world which has been represented by not just complete academic dictionaries and grammar references, but also large and well-structured national corpora. Needless to say, the Russian National Corpus challenges not only researches and practitioners of the Russian language and starts "the parade of aims", according to Shmelyova (Shmelyova. 2007), but also those professionals who work in the field of translation and interpreting studies and integrate corpora into translator education. Another reason explaining why electronic corpora are still not used widely in the language classrooms, among others by translators either, is probably because they have not been exposed to the potential of corpus during their own education.

For reference, regular CULT (Corpus Use and Learning to Translate) conferences have become a European tradition since 1997. They significantly contribute to corpus linguistics, corpus-based translation studies, language teaching and translator training. Corpora have proved to be very useful when students (as well as other trainees) have to master their skills, compensate for insecurities in the target language and culture.

What is language corpus

How do modern linguistics define language corpus? Theoretically, any collection of more than one text can be called a corpus: *corpus* (Latin plural *corpora*, English plural *corpuses* or, commonly, *corpora*) is Latin for *body*). The term is used to mean a number of rather different things. It may refer to any collection of linguistic data (written, spoken, or a mixture of the two), although many practitioners prefer to reserve it for collections which have been organized or collected with a certain purpose in view, generally to characterize a particular state or variety of one or more languages. *"Oxford Advanced Learner's Dictionary of Current English"* defines corpus as a collection of written and spoken texts. Laviosa sees corpus is "a collection of authentic texts held in electronic form and assembled according to specific design criteria" (Laviosa, 2010: 80). In *"Explanatory Translator's Dictionary",* Nelubin defines the term of corpus as follows: "1. Exemplary collection of utterances, selected for linguistic analysis and introduced as written or recorded texts. 2. Entire reference system of language products created by language-users" (Nelubin, 2006: 94).

The Russian National Corpus itself sets out corpus as a reference system based on an electronic collection of texts composed in a certain language. A national corpus represents that language at a stage (or several stages) of its development in all the variety of genres, styles, territorial and social variants of usage, etc.

In most cases, national corpora are created by linguists for academic research and language teaching. Most of the major world languages have their own corpora. A well-recognized example is the British National Corpus (BNC), a huge corpus of 100 million words, which is used as a model for many modern corpora. Some other popular English language corpora include:

- The Brown Corpus: A corpus of written American English from 1961. Compiled at Brown University
- The LOB Corpus: (Full name: The Lancaster-Oslo-Bergen Corpus) A corpus of written British English from 1961

- The London-Lund Corpus: A corpus of spoken British English from the 1960s and early 70s. Recorded and transcribed at The Survey of English Usage (University College, London) and Lund University (Sweden)
- The Helsinki Corpus: A diachronic corpus consisting of a selection of texts covering the Old, Middle, and Early Modern English periods
- Australian Corpus of English (ACE)
- Wellington Corpus (New Zealand)
- The International Corpus of English – East African component
- Lancaster/IBM Spoken English Corpus (SEC)
- Wellington Spoken Corpus (New Zealand)
- Corpus of Early English Correspondance,
- Innsbruck Computer-Archive of Machine-Readable English Texts (ICAMET)

The Russian National Corpus (RNC) includes primarily original prose representing standard Russian (from the middle of the 18th century) but also, albeit in smaller volumes, translated works (parallel corpora) and poetry, as well as texts, representing the non-standard forms of modern Russian: spoken (recordings of oral speech, spontaneous and public) and dialectal. *Parallel corpora*, which contain aligned texts from the learners' native language and translations in the target language (or vice versa) can significantly enrich the translation study environment.

Laviosa (Laviosa, 2010: 80-81) classifies corpora according to six sets of contrastive parameters:

1) *Sample (finite) or monitor (open)*

A sample (or finite) corpus contains abridged or full texts that have been gathered to represent a language or language variety. A monitor (open) corpus is supplemented with new texts and keeps increasing in size.

2) *Synchronic or diachronic*

A synchronic corpus consists of texts produced at one particular time, while a diachronic corpus is made up of texts produced over a long period of time.

3) *General (reference) or specialized*

A general (or reference) corpus represents a language for every day, general usage. A specialized corpus represents a language for special purposes, i.e. ESP – English used in specialized field of knowledge.

4) *Monolingual, bilingual or multilingual*

These are corpora containing texts produced in a single language or in two or more than two languages accordingly.

5) *Written, spoken, mixed (written and spoken) or multi-modal*

These corpora consist of written or/ and recorded spoken texts; or texts produced by using a combination of various semiotic models, e.g., language, image or sound for a multi-modal corpus.

6) *Annotated or non-annotated*

An annotated corpus contains textual or contextual information and/or interpretative linguistic analysis added to raw material data. Corpora can be annotated at different levels of linguistic analysis: phonological, morphological, semantic, parts of speech, lexical, syntactic, discourse, pragmatic, or stylistic. A non-annotated corpus contains plain text that has not been analysed.

Short insight into translation competence

Integrating corpora into the translation classroom is neither a fashionable trend, nor anything of debated translator's "musts". Although Stewart ironically points out that today translators "should be culture-aware, function-aware, register-aware, frequency-aware, ever alert to context and purpose, to co-text, to source language and target language conventions, requirements and restraints" (Stewart, 2009: 29), corpora have been increasingly used in descriptive studies of translation, translator training, translation quality assessment, and computer-added translation. The introduction of corpora in translation classroom was put forward by Mona Baker, a renowned professor of Translation Studies and Director of the Centre for Translation and International Studies at the University of Manchester in England, in 1993. Since then, competent use of electronic text corpora in conjunction with corpus analysis tools help teach better language service providers by enhancing both the quality of translators' work and their productivity. Corpora is not simply about explaining how these tools work or using them to translate. It is mostly about developing *translation competence.* Translation competence is a complex concept that has been addressed by a number of researchers in the field of Translation Studies. The more complete and coherent definition belongs to PACTE (*Process of the Acquisition of Translation Competence and Evaluation)* research group from Barcelona (*the Universitat Autònoma de Barcelona*) as it includes all the aptitudes and skills needed to translate. Their definition is the following: "Translation competence is the ability to carry out the transfer process from the comprehension of the source text to the re-expression of the target text, taking into account the purpose of the translation and the characteristics of the target-text readers".

The insight into translation competence correlates with one of the challenges of the Bologna reforms reflected in a credit system in terms of students' activities and the competences that they acquire. In the case of translation competence, the required sub-competences involve mainly procedural rather than declarative knowledge, learning to use strategies and methodologies. The teacher is no longer the sole source of information and authority. Correspondingly, corpus methodology reinforces autonomy and responsibility.

To be competitive in the 21st century, translators and interpreters cannot do without information technology skills. According to the Third-Generation Educational Standards of the Russian Federation, modern translators-university graduates must possess a number of professional translation sub-competences, including ability to use modern educational and information technologies for upgrading their professional qualifications and searching for professional information in both paper and electronic sources, including corpora.

The development of the professional sub-competence to use electronic corpora to translate contributes a lot to the wider construct of translation competence. The synergies between electronic corpora and other instructional resources can be exploited to maximum pedagogical effect. Today, different types of electronic corpora are used in translation teaching, with emphasis on those that, rather than simply furnishing ready-made solutions, encourage students' critical reflection.

Corpora in the translation classroom

According to some western scholars (Leech, Grabovsky) the conspicuous convergence between teaching and learning corpora could be considered at three interrelated levels, namely (1) direct use of corpora in teaching (teaching about language corpora, teaching to browse the corpus, and browsing the corpus to teach; (2) indirect use of corpora (development of language teaching materials, development of testing and assessment tools; development of supplementary language teaching materials and mini-lexicons); (3) teaching-oriented corpus-development (compiling learners' corpora, compiling parallel corpora, etc.).

How might the enrichment of the translation study environment be achieved?

The least controversial issue is that corpora are a *source of materials* for translator training. It could be taken into consideration by teachers who start working in this field. The corpora are selected and controlled by the teacher to provide real-life examples and exercises. Depending on the nature of the task, the students' learning can be deductive or inductive, and students see that apart from the teacher's knowledge or 'intuition' there are other sources of authority. Here Marco & Lawick differentiate *corpus-based* and *corpus-driven* learning. In the first case, teachers select material from corpora to design classroom materials for specific objective: "the theory precedes the data, and the data are mainly used in support of the theory". In the second case, students have access to an enormous range of language data and they have to learn how to use this data for autonomous learning: "the corpus is seen as more than a repository of examples to back pre-existing theories or a probabilistic extention to an already well defined system" (Marco & Lawick, 2009: 9-11). Almost the same idea is expressed in the approach termed *Data Driven Learning* (Millar & Lehtinen, 2008: 62), when students can act as 'researchers' in the classroom by using authentic corpus data to identify language patterns.

Some examples of corpus-related activities are translation tasks designed within a task-based methodology, which, in its turn, is rooted in the communicative approach. The following four kinds of tasks are envisaged for translator training: *cloze texts*, *multiple choice exercises*, *translation of short passages yielded by a concordancer* and *concordance analysis*. (A concordancer is a computer programme that automatically constructs a concordance. Concordancers are used in corpus linguistics to retrieve alphabetically or otherwise sorted lists of linguistic data from the corpus in question, which the corpus linguist then analyses.) Within the course, it makes sense to combine corpus-based and corpus-driven work, so that the student can gradually shift from the former to the latter as translation skills are not developed overnight.

Corpora allow preparing classroom materials designed to raise awareness about more complex phenomena, like *semantic prosody* or *explicitation* as a translation universal. The term 'semantic prosody' has been based upon a parallel with discussion of prosody in phonological terms. Semantic prosody has recently aroused considerable attention within corpus linguistics. Stewart points out that many uses of words and phrases tend to occur in a certain semantic environment, for example "the word *happen* is associated with unpleasant things – accidents and alike" (Stewart, 2009: 31) If we wish to translate the sentence 'She sat through the opera', we should be aware that the expression *SIT through (something)* is associated with a prosody of boredom or discomfort (ibid: 29) The methodology of translation semantic prosody in corpus linguistics has been developed by a number of scholars and practitioners in some European universities (University of Macerata, Italy, and others) and opens new opportunities for translators.

As one of the universals of translation, *explicitation* is the process of rendering implicit information in the target text. Corpus technology sheds some light on the complex relationship between translation, text length and explicitation. An awareness of what makes translation longer (or shorter) and more explicit than a source text can help translation teachers make more informed decisions about the translation process. Explicitation can be regarded as an important component of translator education. Explicitation is obligatory when the grammar of the target language "forces the translator to add information which is not present in the source text, but can occur voluntary when, for no grammatically compelling reason, translators distance themselves from the source text in a way that makes the target text easier to comprehend" (Frankenberg-Garcia, 2009: 48).

Another opportunity language corpora give to students is creating their own *DIY (do-it-yourself) corpus*. It is a corpus of texts that put together for the sole purpose of providing information – factual, linguistic or field-specific – for the purposes of completing a translation task. Corpora constructed for the specific purpose of being use as translation recourse for the specific translation task have also been called *ad hoc corpora* or *disposable corpora* (Sánchez-Gijón, 2009), or *a local learner corpus* – a term put forward by Seidlhofer. Corpora are open, i.e. texts may be constantly added (and some texts may be removed) to reflect the fact that concepts and terms within certain fields are regularly evolving. The translator needs to learn to construct such corpora, since by now ready-made specialised electronic text corpora are few and far between.

In the translation classroom, electronic corpora could also serve as a useful analysis tool in various aspects:

- confirming intuitive decisions;
- verifying or rejecting decisions based on other tools such as dictionaries;
- obtaining information about collocates;
- reinforcing knowledge of target language patterns;
- learning how to use new expressions.

Conclusion

Despite achievements and enthusiasm within academic settings, several challenges can be identified for translation education.

On the one hand, translator-oriented e-learning materials have to be provided so as to reach those professionals who are eager to learn about corpora and with corpora. These materials should be contrastive in focus (i.e., why/ when use corpora instead of the Web/ dictionaries?), and include substantial practice primarily with those tools and facilities that translators (rather than linguists or language learners) are likely to find of immediate relevance (e.g. concordancing should be given priority over word-listing). Such practice should be embedded in translation-relevant tasks and should not neglect serendipitous turns encouraging the exploration of language and translation issues. On the other hand, corpus-construction and corpus-searching should be ideally integrated with CAT tools, so as to reach the largest possible number of professionals, including less technologically enthusiastic.

References

1. A. Beeby, et al. Investigating Translation Competence: Conceptual and Methodological Issues. *Translators' Journal*, 50 (2) (2005), 609-619.

2. A. Frankenberg-Garcia, Are Translations Longer than Source Texts? A Corpus-Based Study of Explicitation, in *Corpus Use and Translating,* ed. by A.Belly, P.R.Ines and P.Sanchez-Gijon (Amsterdam/Philadelphia: John Benjamin Publishing Company, 2009), 47-58.

3. S. Laviosa, Corpora, in *Handbook on Translation Studies,* ed. by Y.Gambier & L.van Doorslaer (Amsterdam/Philadelphia: John Benjamin Publishing Company, 2010), 80-86.

4. J. Marco & H. van Lawick, Using Corpora and Retrieval Software as a ource of Materials for the Translation Classroom, in *Corpus Use and Translating,* ed. by A.Belly, P.R.Ines and P.Sanchez-Gijon (Amsterdam/Philadelphia: John Benjamin Publishing Company, 2009), 9-28.

5. T. McEnery & A. Wilson, *Corpus Linguistics* (UK: Edinburgh University Press, 2001)

6. T. Millar & B. Lehtinen, DIY Local Learner Corpora: Bridging Gaps Between Theory and Practice, in *The JALT CALL Journal*, 4 (2) (2008), 61-72.

7. L. Nelubin, *Explanatory Translator's Dictionary* (Moscow: Publishing House Flinta, 2006), in Russian.

8. *The Russian National Corpus and the Problems of Humanitarian Education* (Moscow: Publishing House TEIS, 2007), in Russian.

9. P. Sánchez-Gijón, Developing Documentation Skills to Build Do-It-Yourself Corpora in the Specialised Translation Course. *Corpus Use and Translating,* ed. by A.Belly, P.R.Ines and P.Sanchez-Gijon (Amsterdam/Philadelphia: John Benjamin Publishing Company, 2009), 109-128.

10. T. Shmelyova, The Corpus Problem Book, in *The Russian National Corpus and the Problems of Humanitarian Education* (Moscow: Publishing House TEIS, 2007), 25-34, in Russian.

11. D. Stewart, Safeguarding and Lexicogrammatical Environment: Translating Semantic Prosody, in *Corpus Use and Translating,* ed. by A.Belly, P.R.Ines and P.Sanchez-Gijon (Amsterdam/Philadelphia: John Benjamin Publishing Company, 2009), 29-46.

Корпус языка для обучения переводу, или Dies diem docet (день учит день)

В.А. Кононова
Сибирский федеральный университет
Россия, 660041, Красноярск, пр. Свободный, 79

Статья предлагает к обсуждению некоторые возможности использования корпуса языка при обучении переводу студентов. Корпус является бесценным источником для студентов-переводчиков, а также солидной базой для развития их профессиональной компетенции. В Российской Федерации, где Национальный Корпус языка появился недавно – в апреле 2004 г. – вопросы активного использования и пополнения этого богатейшего источника звучат особенно актуально.

Ключевые слова: корпус языка, Национальный корпус русского языка, профессиональная компетенция переводчика, задания с использованием корпуса.

Научная специальность: 10.00.00 – филологические науки.