

The Use of Corpora and Corpus Technology in Academic and Professional Environment

استخدام المدونات وتكنولوجية المدونات في الأوساط المهنية والأكاديمية

إلهام بزاوشة Ilhem BEZZAOUCHA

University of Algiers 2, Algeria

ilhem.bezzaoucha@univ-alger2.dz

DOI: 10.46314/1704-021-001-017

Received date: 10/05/2021

Revised date: 04/06/2021

Publication date: 20/07/2021

Abstract:

The use of corpora and concordancers in translation studies has grown exponentially since the mid 1990s' with a plethora of literature advocating their use and benefits in both professional and academic settings. In translator training, efforts are being made to incorporate the use of corpora in masters' programmes and to offer modules on corpora for translation as the use of translation memory (TM) systems and Computer-Aided Translation (CAT tools) dominate the translation profession. This paper sets out to demonstrate the usefulness of corpora and corpus technology at the different stages of the translation process as it has now become of paramount importance to assess the extent to which corpora are of added value for translation quality in both professional and academic environments.

Keywords: Corpus; Corpora; Corpus technology; Translator training; Professional translator; Translation technology.

ملخص:

عرف استخدام المدونات في دراسات الترجمة تطورا كبيرا منذ منتصف تسعينيات القرن الماضي مع عدد كبير من الأدبيات التي دعت إلى استخدامها وتنبأت بفوائدها في الأوساط المهنية والأكاديمية. تُبذل الجهود في تكوين المترجمين بإدماج استخدام المدونة في برامج الماجستير واقتراح مواد في الترجمة حول المدونات، حيث إن استخدام أنظمة ذاكرة الترجمة والترجمة المدعمة بالحاسوب (أدواتها) يهيمن على مهنة الترجمة. تهدف هذه الورقة البحثية إلى توضيح فائدة المدونات وتكنولوجيا المدونات في المراحل المختلفة

من عملية الترجمة، حيث أصبحت الآن ذات أهمية قصوى لتقييم مدى القيمة المضافة لجودة الترجمة في كل من الأوساط المهنية والأكاديمية.

الكلمات المفاتيح: مدونة: تكنولوجيا المدونات: تكوين المترجمين: مترجم محترف: تكنولوجيا الترجمة.

1. INTRODUCTION

The use of corpora and corpus technology for translation purposes has been advocated since the nineties. Corpora and concordancers in the classroom have the ability to raise students' awareness of the intricacies of the languages as they provide students with authentic linguistic data and enhance their capabilities by suggesting accurate and idiomatic words and phrases most unlikely to be found in traditional resources; corpus linguistics, as a methodology which focus on the identification of recurrent patterns of linguistic behaviour in real linguistic data, provides the appropriate tool to test hypotheses about norms and regularities in translated texts.

Regularities of translators are individual linguistic habits manifested through consistently different (unconscious) patterns of choices, independently of the source texts.

Corpora have thus become an added value for translation quality in both professional and academic environments

2. Corpora and Concordancers for Trainee translators

The introduction of corpora in Translation Studies was put forward by Mona Baker (1995) in her seminal article entitled “*Corpus linguistics and Translation Studies: Implications and Applications*”. Since then, the role of corpora has grown increasingly important in both corpus-based translation studies and applied corpus-based translation studies. While corpus-based methodology serves for identifying the distinctive features of the language of translation such as “the specific constraints, pressures, and motivations that influence the act of translating and underlies its unique language” (Baker, 1998 480), applied corpus-based translation studies

considers the use of corpora and concordancers for translation teaching. Broadly speaking, there are two major complementary approaches to using corpora and corpus technology for translation teaching known as “Corpus use for learning to translate” and “Learning corpus use to translate” (Beeby et al, 2009): teachers can prepare learning materials and tasks using corpora and students will become autonomous users of corpora as part of their translation aptitude.

Domain-specific corpora in other languages than English are not as easily retrieved, especially not packaged with a set of corpus analysis tools. The notion of “disposable” or “Do-It-Yourself” (henceforth DIY) corpora has been put forward as a corpus that translators and other linguists would build themselves to rapidly search for information on the spot.

In this DIY approach which involves constructing one’s own corpus, students solve specific translation problems by finding the most accurate words, collocations and phrases: “we can regard them [disposable, ad hoc corpora] as performance-enhancing tools in translation or, more precisely, as decision-making tools for lexical and textual knowledge management in translation.” (Varantola, 2003: 59). Hence, corpora and corpus interrogation tools serve as documentation tools of sorts.

Norms are seen as governing the activity of translation at all levels, from the decision to translate a text in the first place to the choice of the strategies implemented in the process of translation and which determine the actual linguistic composition of a translation (Zanettin, 2014)

The use of corpora and concordancers for translation teaching purposes is widely documented in strengthening translation norms. Corpora fall into two main categories: 1) comparable corpora defined as “a collection of texts composed independently in the respective languages and put together on the basis of similarity of content, domain and communicative function.” (Zanettin, 1998: 614), and 2) parallel corpora consisting of original texts and their translations. Parallel corpora may provide a wider inventory of possibilities than a single translator is likely to come up with.

The visual output and functionality of corpus analysis software, most frequently referred to under the umbrella-term 'concordancer', has helped students to uncover patterns in language usage that had been hidden from view. However, the concordancer only acts as a facilitator; the processing and interpretation of the data depend on "the observational and generalizing skills of the investigator" (Leech 1992: 114).

3. The Use of Corpora and Corpus Technology in the Professional Environment

Merely equipped with dictionaries, glossaries and alike, translators and interpreters used to practice their work on a daily basis. Nowadays, As Bowker and Corpas Pastor (2015) state: "In today's market, the use of technology by translators is no longer a luxury but a necessity if they are to meet rising market demands for the quick delivery of high quality texts in many languages."

However, there seems to be little or no demand from the translation market requiring translators to use corpora and concordancers in particular. Actually, there is no particular pressure from the market to train translators to use corpora (Frankenberg-Garcia, 2015) which may turn "counterproductive" for teachers involved in the teaching of trainee translators. Professional translators had little explicit knowledge of corpora as a recent research conducted by Frankenberg-Garcia (2015) in two international translator forums, Proz.com and Translator's Cafe, reveals that there are no references to corpora when compared to several daily queries about translation-memory systems and CAT tools. Another significant reason why corpora are underestimated as sources of documentation (terminology, phraseology and knowledge rich content) is their limited or non-existent availability for many specialised subjects and language pairs (Kübler, 2011: 66).

Caratalla puertas (2015) describes the relationship between corpora and professional translators as omnipresent and invisible as professional translators actually resort to corpora as they look for parallel texts using

search engines or search previous translations. Google is used as a “raw” mega-corpus; the largest ever and the broadest in scope (Borja, 2008).

It is considered adequate to extract domain-specific terminology. In the lack of printed specialised dictionaries and bilingual parallel texts for many specific subjects, the monolingual texts retrieved from the Internet constitute a direct source for reliable consultation.

4. The added value of corpora in the translation process

While corpora help enhance students’ translations by providing information missing from dictionaries, especially regarding term choice and idiomatic expressions, corpora are of added value for the translation process. When it came to assessing the usefulness of corpora on the translation process, Kubler et al (2015) investigated the efficiency of corpus use for students during terminology processing and LSP translation in the field of earth sciences and found that “besides an obvious and expected positive outcome on the validation and translation of terms, there is an interesting positive influence of corpora in the translation process on elements other than those related to terminology, such as collocations and various genre- and discourse-based features”.

Parallel corpora are, for example, repertoires of strategies used by past translators, as well as repertoires of readily-made translation equivalents.

Professional translators as well as trainee translators are able to investigate the information contained in corpora using corpus analysis tools, they provide all the frequency count of the search (query) word or phrase in the corpus, then by clicking on the word in the display screen, all concordance lines in which the search word or phrase appear will be shown. The "concordancer" displays information in the screen together with a span of co-text to the left and right.

Corpora are not meant to replace other resources. Rather they are valuable reference tools which- along with other resources- have a definite place in the translation process.

4.1. Collocations

Firth, in a quite famous quote, states that “you shall know a word by the company it keeps” (Firth 1957: 11). Words are always embedded in contexts, in particular semantic fields. Relying on dictionaries to investigate the usage of a word provides a set of diversified equivalents accompanied by potential collocates. Such a treatment is quite reductive, and translations are generally not exemplified in real contexts. Most dictionaries fall short of offering examples which can go far beyond the scope of prototypical meanings along with their collocates.

The occurrences of a word are categorised in their respective contexts to examine the word’s collocational behaviour based on the different translations. The final step consists in the capture of the regularities of the word in its collocational environment.

Several teaching activities revolve around corpus querying in order to identify equivalent terms in the target language and ways to couch terms in phraseology that is acceptable for the target language text.

4.2. Verifying or rejecting decisions taken based on corpora

We will see how knowledge of the semantic and pragmatic behaviour of a *word* is better informed when the item is looked at in real contexts and through its different translations into other languages. Contexts allow translators to search through corpora to find information and spot linguistic patterns that might help them to complete a new translation. The inclusion of formal professionals’ output is intentional as to show how things work in real-life situations, since those trainee-translators/interpreters have no access to stored, automatic solutions.

Various types of corpora can be assessed to settle on one decision: DIY corpora (monolingual, comparable, parallel corpora), online corpora (BNC or the COCA for the English language), large online translation databases (Linguee, Tradooit or Reverso Context) as well as Google search. Finally, quality corpora can help students find better solutions than some translation memories and bilingual resources, which tend to provide ready-made translations that, are not always suitable.

5. Analysing and exploiting corpus data in the classroom

According to Kübler, Mestivier and Pecman (2018) the most frequent errors that students manage to avoid using corpora are: term translated by a non-term, incorrect collocation, incorrect choice and wrong preposition. An “incorrect collocation”, for instance, is one of the most frequent translation mistakes that students make and one that could be easily avoided by using corpora. In this vein, they believe “that taking the students through a full process of term identification, the associated terminological and phraseological analysis, and finally translation, along with a step-by-step demonstration of how to use the concordancer and how the corpus can be used to answer each question, whenever possible, will furnish them with the necessary tools for their future careers as translators (Kübler, Mestivier and Pecman, 2018: 818).

6. A case study: Pride and Prejudice

Fiction texts have only rarely been analysed by corpus linguistic techniques. Little importance has been given to them as the literature suggests. Women, for instance, are frequently described in terms of their physical appearance, as in shown by a high frequency of adjectives such as “beautiful”, “pretty” and “lovely” as a collocation. References to men are usually modified by adjectives related to status such “important”, “great” and “prominent”. This is the result of the patriarchal vein that still runs in the background of our societies. Here’s a series of adjectives taken from Jane Austen’s novel, “Pride and Prejudice”. The corpus-driven approach direct our attention to high-degree words that mirror the lexical items

chosen strategically or unconsciously for hinting at meanings between the lines.

While “beautiful” and “lovely” still pertain to women, ‘pretty’ can be used also as an intensifier.

360 : **h** Oh! She is the most [beautiful](#) creature I ever beheld! But there is one of [402](#) : Bingley thought her quite [beautiful](#), and danced with her twice. Only think of [552](#) : not conceive an angel more [beautiful](#). Darcy, on the contrary, had seen a collection

[796](#) : intelligent by the [beautiful](#) expression of her dark eyes. To this

[2016](#) : could do justice to those [beautiful](#) eyes?**h** **g**It would not be easy, indeed, [12946](#) : sure you could not be so [beautiful](#) for nothing! I remember, as soon as ever I saw

- | | | |
|------------------------|------------------------------|---|
| 2700 : | them. -- Miss Bennet's | lovely face confirmed his views, and established all |
| 5765 : | to see it healthful and | lovely as ever. On the stairs were a troop of little |
| 9952 : | finding her otherwise than | lovely and amiable. When Darcy returned to the |
| 355 : | of them, you see, uncommonly | pretty . h g You are dancing with the only handsome |
| 362 : | behind you, who is very | pretty , and I dare say very agreeable. Do let me ask |
| 463 : | were about five times as | pretty as every other women in the room. No thanks |
| 556 : | He acknowledged to be | pretty , but she smiled too much. Mrs. Hurst and |
| 610 : | there were a great many | pretty women in the room, and which he thought the |
| 791 : | scarcely allowed her to be | pretty ; he had looked at her without admiration at |

- [961](#) : fine eyes in the face of a [pretty](#) woman can bestow. **h Miss Bingley**
- [1625](#) : and when I am in town it is [pretty](#) much the same. They have each their
- [1679](#) : some verses on her, and very [pretty](#) they were. **h g And so ended his**
- [5337](#) : easily falls in love with a [pretty](#) girl for a few weeks, and when accident
- [6179](#) : Collins, was a very genteel, [pretty](#) kind of girl. She asked her at different
- [6485](#) : and Mrs. Collins's [pretty](#) friend had moreover caught his fancy very much.
- [6562](#) : will give you a very [pretty](#) notion of me, and teach you not to believe a
- [6943](#) : she had somehow or other got [pretty](#) near the truth. She directly replied,
- [8142](#) : I do not think it is very [pretty](#); but I thought I might as well buy it as not.
- [8355](#) : it has been shifting about [pretty](#) much. For my part, I am inclined to believe
- [8685](#) : her return, agitation was [pretty](#) well over; the agitations of former partiality
- [9146](#) : thought this was going [pretty](#) far; and she listened with increasing
- [9203](#) : were shewn into a very [pretty](#) sitting-room, lately fitted up with greater
- [9991](#) : you thought her rather [pretty](#) at one time. **h g Yes, h replied Darcy,**
- [11965](#) : or Jane at most. You know [pretty](#) well, I suppose, what has been done for the
- [12202](#) : handsome, and said many [pretty](#) things. **g He is as fine a**

fellow, **h** said

[12695](#) : -- and her nieces are very

[pretty](#) behaved girls, and not at all handsome: I like

[13849](#) : she found that it had been

[pretty](#) much the case. **g** On the evening before my

However, the adjective ‘handsome’ which is nowadays attached to men, has been applied to women and men alike:

[99](#) : am sure she is not half so [handsome](#) as Jane, nor half so good humoured as Lydia.

[261](#) : was quite young, wonderfully [handsome](#), extremely agreeable, and, to crown the whole,

[357](#) : are dancing with the only [handsome](#) girl in the room, **h** said Mr. Darcy, looking

[367](#) : **g** she is tolerable; but not [handsome](#) enough to tempt me; and I am in no

[420](#) : him. He is so excessively [handsome](#)! and his sisters are charming women. I never...

It can be said that concordance investigation is useful in helping stylistic analysts select linguistic features for a detailed qualitative analysis in a more objective fashion. Stylisticians can become more objective when choosing fragments of texts that contain some important features relatable to some interpretative issues.

7. CONCLUSION

The literature seems to show that in regards to teaching environments, there are only a few curricula that include the use of corpora in translation teaching, as opposed to translation memories, which are a must for the labour market.

In this article it was tried to show how a corpus analysis tools such as concordancers can be a useful performance-enhancing aid in translating, they are likely to help both trainee translators and professional translators to

improve the quality of their translations and their productivity, especially when translating special field texts into a foreign language. Informed by the proliferation of CAT tools in translation training, corpora have somehow emerged to offer platforms to enhance students' learning experience.

Current teaching trends in classrooms aspire to make translating assignments more reality/professional-like, trainers aim to put trainees in authentic scenarios that are likely to instil the general learning atmosphere with a sense of general satisfaction. Technology-driven activities for didactic purposes seem to foster learners' autonomy and sharpen their competitive edge in evaluating and rethinking a set of finite possible solution to arising translation problems.

Corpora as part of the new digital revolution can offer trainees with another set of tools to match the various phases of an assignment (from preparation to the final reformulation). Each phase can benefit from a specific type of corpus: monolingual corpora in the target language, parallel corpora in the field of study, learner corpora derived from students' output and video corpora for a real touch of professionalism.

REFERENCES

- Aston, G. (1999) "Corpus use and learning to translate." *Textus* 12: 289-314.
- Baker, Mona. (1995) "Corpora in translation studies: An overview and suggestions for future research." *Target* 72: 223-244.
- Baker, Mona. (1998) "Réexplorer la langue de la traduction: une approche par corpus." *Meta* 43.4: 480-485.
- Beeby, A., Rodríguez-Inés, P. & Sánchez-Gijón, P. (2009) "Introduction." *Corpus Use and Translating*. Ed. A. Beeby, P. Rodríguez Inés and P. Sánchez-Gijón. Amsterdam: John Benjamins. 1-8.
- Bernardini, S., Castagnoli, S. (2008) "Corpora for translator education and translation practice." *Topics in Language Resources for Translation and Localisation*. Ed. E. Yuste Rodrigo. Amsterdam/Philadelphia: John Benjamins. 39-55.

- Bernardini, S. (2015) “Exploratory learning in the translation/language classroom: Corpora as learning aids.” *CULT Conference, 26-29 May*, Alicante.
 - Bowker, L. (1998) “Using Specialized Monolingual Native-Language Corpora as a Translation Resource: a Pilot Study.” *Meta* 43.4: 631-651.
 - Bowker, L. (1999) “Exploiting the potential of corpora for raising language awareness in student translators.” *Language Awareness* 8: 160-73.
- Bowker, L. (2002) *Computer-Aided Translation Technology: A Practical Introduction*. University of Ottawa: Ottawa Press, Didactic of Translation Series.
- Bowker, L., Pearson, J. (2002) *Working with Specialized Text: A Practical Guide to Using Corpora*. UK: Routledge.
 - Bowker, L., Barlow, M. (2008) “A comparative evaluation of bilingual concordancers and translation memory systems.” *Topics for Language Resources in Translation and Localisation*. Ed. E. Yuste Rodrigo. Amsterdam/Philadelphia: John Benjamins. 1-22.
 - Bowker, L., Corpas Pastor, G. (2015) “Translation Technology.” *Handbook of Computational Linguistics*. Ed. R. Mitkov. Oxford: Oxford University Press.
 - Frankenberg-Garcia, A. (2005) “A Peek into What Today’s Language Learners as Researchers Actually Do.” *International Journal of Lexicography* 18.3: 335-355. [Links]
 - Frankenberg-Garcia, A. (2015) “Training translators to use corpora hands-on: challenges and reactions by a group of 13 students at a UK university.” *Corpora* 10.2.
 - Fernandes, Lincoln.-P. (2007) “On the Use of a Portuguese-English Parallel Corpus of Children’s Fantasy Literature in Translator Education.” *Cadernos de Tradução* 20: 141-163.
 - Fulford, H., Granell-Zafra, J. (2005) “Translation and Technology: a Study of UK Freelance Translators.” *The Journal of Specialised Translation* 4: 2-17.

- Gallego-Hernández, D. (2015) “The use of Corpora as translation resources: a study based on a survey of Spanish professional translators.” *Perspectives: Studies in Translatology* 23.3: 375-391.
- Garcia, I. (2009) “Beyond Translation Memory: Computers and the Professional Translator.” *The Journal of Specialised Translation* 12: 199-214. [Links]
- Kenny, D. “Translation memories and parallel corpora: Challenges for the translator trainer.” *Across Boundaries: International Perspectives on Translation Studies*. Ed. D.
- Krüger, R. (2012) “Working with corpora in the translation classroom.” *Studies in second language learning and teaching* 4: 505–525. [Links]
- Kübler, N. (2003) “Corpora and LSP translation.” *Corpora in Translator Education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St Jerome. 25-42. [Links]
- Kübler, N. (2011) “Working with Corpora for Translation Teaching in a French-speaking setting”. *New Trends in Corpora and Language Learning*. Ed. A. Frankenberg-Garcia, L. Flowerdew, and G. Aston. London: Bloomsbury, 62-80. [Links]
- Laviosa, S. (2010) “Corpora.” *Handbook of Translation Studies*. Y. Gambier, L. van Doorslaer, 1. 80-86.
- Maia, B. (2003) “Some languages are more equal than others. Training translators in terminology and information retrieval using comparable and parallel corpora.” *Corpora in Translator Education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St Jerome. 43-53.
- Marco Borillo, J., Van Lawick, H. (2009) “Using corpora and retrieval software as a source of materials for the translation classroom.” *Corpus Use and Translating*. Ed. A. Beeby, P. Rodríguez-Inés and P. Sánchez-Gijón. Amsterdam/Philadelphia: John Benjamins. 9-28.
- Olohan, M. (2004) *Introducing Corpora in Translation Studies*. Routledge.
- Pearson, J. (1996) “Electronic texts and concordances in the translation classroom.” *Teanga* 16: 86-96.

- Pearson, J. (2000) “Teaching terminology using electronic resources.” *Multilingual Corpora in Teaching and Research*. Ed. S.P. Botley, A.M. McEnery, and A.Wilson. 92-105.
- Pearson, J. (2003) “Using Parallel Texts in Translation Training Environment.” *Corpora in Translator education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St. Jerome. 15-24.
- Peters, C., Picchi, E. (1998) “Bilingual reference corpora for translators and translation studies.” *Unity in Diversity? Current Trends in Translation Studies*. Ed: L. Bowker, M. Cronin, D. Kenny and J. Pearson. Manchester: St. Jerome. 91-100.
- Picton, A., Fontanet, M., Pulitano, D., Maradan, M. (2015) “Corpora in Translation: addressing the Gap between the Scholars’ and the Translators’ Point of View.” *CULT Conference, 26-29 May*, Alicante.
- Rodríguez-Inès, P. (2009) “Evaluating the process and not just the product when using corpora in translator education.” *Corpus Use and Translating*. Ed. A. Beeby, P. Rodríguez-Inés and P. Sánchez-Gijón. Amsterdam: John Benjamins. 129-149.
- Rossi, C., Frérot, C., Falaise, A. (2016) “Integrating controlled corpus data in the classroom: a case-study of English NPs for French students in specialised translation.” *Corpus-based studies on language varieties*, Linguistic insights, Bern: Peter Lang.
- Varantola, K. (2003) “Translators and disposable corpora.” *Corpora in Translator Education*. Ed. F. Zanettin, S. Bernardini and D. Stewart. Manchester: St Jerome. 55-70.
- Wilkinson, M. (2005) “Using a Specialized Corpus to Improve Translation Quality.” *Translation Journal* 9.3.
- Williams, I.A. (1996) “A translator’s reference needs: dictionaries or parallel texts.” *Target* 8: 277-299.
- Zanettin, F. (1998) “Bilingual Comparable Corpora and the Training of Translators.” *Meta* 43.4: 613-630.

- Zanettin, F. (2001) “Swimming in words: corpora, translation, and language learning.” *Learning with Corpora*. Ed. G. Aston. Athelstan. 177-197.
- Zanettin, F. (2002) “Corpora in Translation Practice.” *Proceedings of the LREC Workshop. Language Resources for Translation Work and Research*. 10-14.
- Zaretskaya, A. Corpas Pastor, A. Seghiri, M. (2015) “Translators’ requirements for translation technologies: a user survey.” *Proceedings of the AIET17 Conference New Horizons in Translation and Interpreting Studies*.