

# Using monolingual corpora in the target language as a translation aid. The example of collocations from Arabic into English

*Ilhem BEZZAOUCHA,*

*University of Algiers 2*

ملّخص:

في ظل الوتيرة المتسارعة لتخصص المعارف، يتعرض المترجم لطيف هائل من النصوص المتخصصة والعامّة الحاملة للمشاكل المصطلحية المختلفة. وبما أن المصطلحات والكلمات لا تقع فرادى، فهي تقترن بكلمات أخرى لتكون سدم معنوية، كان من اللازم الاعتناء بهذه الكليات اللغوية.

يعنى هذا المقال بأهمية استغلال المتون الأحادية اللغوية كأداة مساعدة للمترجم في سعيه لكتابة نصوص طبيعية في اللغة المنقول إليها. وننوه بأهمية استخدام المتون الأحادية اللغة في اللغة المنقول إليها لتحسين نوعية الأعمال المترجمة.

## **Abstract:**

Faced with the ever-increasing specialization of knowledge, trainee translators have to cope with a wide range of general and specialized texts with varied terminological problems. As terms and words do not come in isolation, they often tend to co-occur with other specific words to form semantic and linguistic nebulas; special attention should be paid to the formation of these linguistic entities.

The objective of this paper is to highlight the importance of monolingual corpora as a support system for the trainee translator in

his quest to produce natural target texts. We argue that trainee translators can substantially improve the quality of their translated works thanks to available monolingual corpora compiled in the target language.

## **Introduction**

Trainee translators are often encouraged to read in their second language to improve their translational competence and performance. Collocations are among the most problematic items in the translation process on which depend the naturalness of the translated text.

Dictionaries, in all their forms, provide useful information about isolated items and concepts; but the lack of a wide-encompassing context is often viewed as a drawback by trainee translators and professionals in general. In recent years, corpora have emerged as additional translation aids to aspiring translators.

## **What is a corpus?**

Roughly speaking, a corpus (plural: corpora) is a compilation of authentic texts in a machine readable form:

“A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent as far as possible, a language or language variety as a source of data for linguistic research.” (Sinclair, 2004: 19)

A random collection of electronic texts cannot constitute a corpus. Corpus design is subjected to rigorous internal and external selection criteria such as: size, content, balance and representativeness.

The British National Corpus (BNC) is often cited as an example of a big representative monolingual corpus in British English. The corpus of Contemporary American English (COCA) is also a large balanced freely-available corpus. This corpus contains 450 million words of various types of texts: newspapers, fiction, spoken, popular magazines and academic texts ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

## **Monolingual corpora and translation**

The seminal work of Baker (1995) laid the foundation of much of the current research in corpus-based translation studies. Corpus-

based translation studies have favoured in their research comparable and parallel corpora for term and phraseology extraction. Such corpora are hard to compile and difficult to access as they are subjected to different copyright restrictions. Monolingual corpora, however, are more pervasive, and contain valuable linguistic information that can benefit both trainee translators and professionals:

“The digital revolution and the wide-spread use of the World Wide Web have proliferated vast amounts of monolingual corpora. Publishing text in one language is much more natural activity than producing parallel texts.” (Koehn and Knight, 2002: 9)

When translation is from the mother tongue into a second foreign language, monolingual corpora will help students to confirm or infirm translational hypotheses, and even put forward a more adequate equivalent that might be totally ignored by students. As Zanettin, Bernardini and Stewart (2014: 6) put it:

“Monolingual corpora can provide information about typical ‘units of meaning’ in the target language or in a specialized subset of it (restricted by topic and/or text type). They can thus help (future) translators opt for natural, ‘native-like’ turns of phrase, appropriate to the communicative situation in which the target text will be operating.”

Bowker (1998) has also found that monolingual corpora improve the overall quality of translated texts in the classroom in respect to subject field understanding, correct term choice and idiomatic usage. Maia (2005) argues that the use of general corpora such as the BNC and COCA allows trainee translators to go beyond the scope of dictionaries as they can observe how words behave in context.

Trainee translators may have intuitions about language use in the target language that will be validated or invalidated by corpus evidence. Even if their intuitions are correct, the language they produce may not be typical or representative of language use. It is worth mentioning that monolingual corpora gather texts entirely produced by native speakers, and language use is expected to be authentic and real.

### **Using collocations from monolingual corpora to enhance naturalness in translated texts**

Monolingual corpora can provide important information about recurrent and frequent linguistic constructions in the target language. These recurrent linguistic data can take the form of a collocation. A collocation is, in a nutshell, a recurrent word combination that appears again and again:

“Collocations are phrases of matching words. They are word combinations that occur and recur frequently in language. They mark which (word) goes with which (word) in language in general. Usually, they are used intuitively and unconsciously by native speakers of a language. Only when they are discussed or disrupted can they be felt by them.” (Ghazala, 2007: 17)

Students of translation often encounter difficulties in finding proper equivalents for certain terms and collocations in the foreign language. This will result in awkward renditions and bizarre word combinations. To counter these problems, one solution would be to provide trainee translators with collection of naturally occurring examples in the subject field written in the target language.

Matching or collocating proper nouns with the proper verbs or other proper adjectives is a constant struggle for trainee translators. When a certain noun occurs with a certain adjective e.g. (انتصار ساحق), students need to attend seriously to the equivalent collocation (smashing victory) to instill the English version with the same spirit. The translator's knowledge of collocations is an important requirement for the overall mastery of the target language (Brashi, 2005: 3). There several types of collocations, the adjective + the noun collocation is certainly among the most recurrent.



**Figure 1. Webcorp interface showing the concordance of the collocation “smashing victory”**

It should be noted that existing monolingual corpora have, despite their relative large size, limitations as a translation aid if the text to be translated gathers recent terminology and phraseology.

Finally, we can only report that this invaluable tool has not been properly exploited by translators in Algeria. This might be due to the fact that trainee translators have yet to be exposed to the potentials of corpus analysis tools in their translation curricula.

## Conclusion

Corpora enable trainee translators to make informed linguistic choices based on language as it is really used in the target language. This practice will increase linguistic awareness in the classroom, and constitutes an added value to the modern practice of translation. Monolingual corpora in the target language will better serve a target-oriented, empirical approach to translation research which in turn will favour an equivalence-oriented scheme based on authentic linguistic data within their contexts of fruition.

Monolingual corpora in the target language are especially interesting for trainee translators as they can view words and their collocations in actual use in the adequate types of text and contexts, and the target text will most likely feel more natural to the target audience.

## Bibliography

1. Baker, Mona. (1995). "Corpora in Translation Studies. An Overview and Suggestions for Future Research", *Target* 7(2): 223-43.
2. Brashi, A. (2005). *Arabic Collocations: Implications for Translation. Published PhD thesis*. Sydney: University of Western Sydney.
3. Bowker, L. (1998). "Using specialized monolingual native-language corpora as a translation resource: a pilot study", *Meta*, 43/4.
4. Ghazala, H. (2007). *Dictionary of Collocations (English-Arabic)*. Beirut: Dar El-ilm.
5. Koehn, Philipp. Knight, Kevin. (2002). *Learning a Translation Lexicon from Monolingual Corpora: Proceedings of the Workshop of the ACL Special Interest group on the Lexicon (SIGLEX)*, Philadelphia, July 2002, pp. 9-16. Association for Computational Linguistics.
6. Maia, B. (2005). "Terminology and Translation-Bringing Research and Professional Training Together through Technology", *Meta: Translator's Journal*, 50. Erudit.
7. Sinclair, Jhon. (2004). "Corpus and Text-Basic Principles". *Developing Linguistic Corpora: a Guide to Good Practic*. Ed. M. Wynne. AHDS Literature, languages and Linguistics, Oxford: Oxford University Computing Services.
8. Zenettin, Federico. Bernardini, Sllvia. Stewart, Dominic. (2014). *Corpora in Translator Education*, Routledge: Newyork.