Using Parallel Corpora for Language Learning

Michael H. Brown, Japan

Michael is a lecturer at Kanda Institute of Foreign Languages in Tokyo, Japan. He also writes about the application of corpus linguistics in language learning, including the use of parallel corpora, at https://corpling4efl.wordpress.com. E-mail: brown-ma@kifl.ac.jp

Menu

Introduction

A common critique of data-driven learning (DDL) and the direct use of corpora in the classroom is that many, if not most, of the corpora available today are inappropriate for language learners because the level of the language in them is too high and the language fragments presented in concordance lines confuse learners. Parallel corpora offer a potential way forward for DDL that addresses these concerns by providing L1 support to learners engaged in DDL activities. This article briefly describes what parallel corpora are and how they can ameliorate certain concerns regarding DDL.

What are parallel corpora?

Although the classroom use of corpora and data-driven learning (DDL) have seen a significant amount of research in recent years, little of the discussion has focused on parallel corpora. A parallel corpus is a collection of texts in one language that is aligned with translations into or from one or more other languages. When concordance lines are produced, lines featuring both the target language and translations from another language are displayed. To illustrate, here are two examples of the concordance display of the English-Japanese Sentence Corpus of Remedial English (SCoRE), from two different web-based interfaces.

Image 1: SCoRE-Corpus Interface (www.score-corpus.org) [3]

Image 2: WebSCoRE Interface (www.antlabsolutions.com/webscore/) [1]

Parallel corpora and language learning

Parallel corpora play a prominent role in domains such as translation studies or contrastive studies of language[6]. However, they have only seen limited use in language learning contexts. There are two general beliefs about corpora that limit their use in language learning contexts. The first belief, assuming that barriers such as lack of electricity or access to the internet are not issues, is that the kind of language found in corpora is too difficult for learners to understand or make use of. And the second one is that even if the language were of the appropriate level, the tools for manipulating and searching corpora are not learner-friendly; they are designed with researchers, not language learners, in mind. Moreover, parallel corpora specifically face issues such as a lack of awareness even among generally corpora-aware instructors, and a lack of accessible tools for searching parallel corpora.
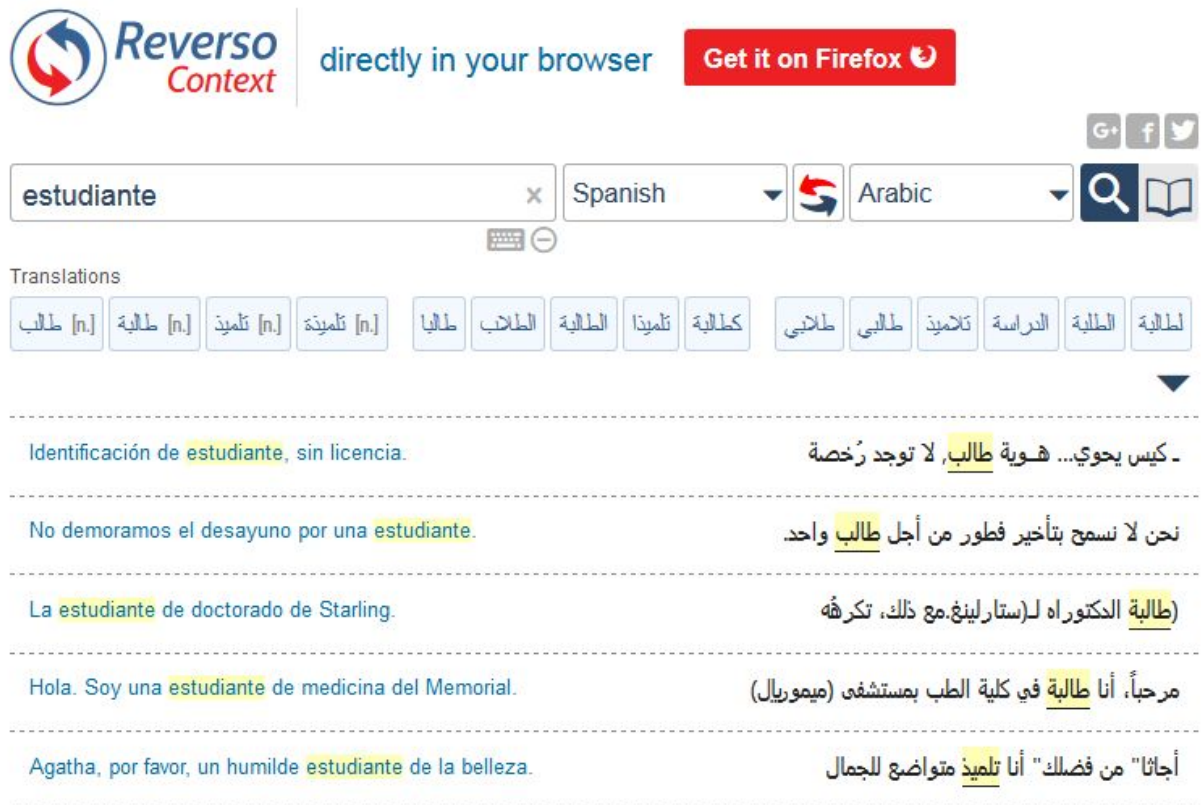
Still, parallel corpora are promising resources which may be able to help instructors and learners overcome some of the barriers to effectively engaging in DDL. The inclusion of L1 concordance lines alongside target language concordance lines can help learners to understand the target language lines [2]. Additionally, parallel corpora can boost L2 comprehension when dictionaries (monolingual or bilingual) and monolingual corpora do not provide enough, or not the right kind of, information sought by learners [4]. Both lexical [10]

and grammatical [2] study by lower proficiency learners can be aided by the use of parallel corpora.

Furthermore, there has recently been an increase in the development of tools and corpora that are designed to be more learner-friendly. The aforementioned English-Japanese SCoRE, for example, is explicitly designed for learners, including the web-based tools for searching it. Another simple to search parallel corpus is the bi-directional Portuguese-English COMPARA [5]. As more parallel corpora featuring more language pairs are developed, we may also see the development of more learner-friendly tools.

There also exist multilingual projects such as Linguee [7], Tatoeba [8], and Reverso Context [9], which produce multi-directional parallel concordance lines. For example, Reverso Context provides sentence pairs into and from 12 languages. The following images illustrate samples of Spanish and Arabic parallel concordance lines, and English and German parallel concordance lines in Reverso Context.

Image 3: Reverso Context, Spanish-Arabic

Image 4: Reverso Context, English-German

Conclusion

As awareness of parallel corpora grows, and more research is conducted using them, we may discover which tasks they are best suited for, which students benefit the most from their use, what pedagogic design principles make them the most useful, and what kinds of tools are the best for helping learners to exploit them.

References

[1] Anthony, L. and Chujo, K. (n.d.) WebSCoRE (website). Available at www.antlabsolutions.com/webscore/

[2] Chujo, K., Kobayashi, Y., Mizumoto, A., & Oghigian, K. (2016). Exploring the Effectiveness of Combined Web-based Corpus Tools for Beginner EFL DDL. Linguistics and Literature Studies, 4(4), 262-274.

[3] Chujo, K., Oghigian, K. and Akasegawa, S. (2015). A corpus and grammatical browsing system for remedial EFL learners. In A. Leńko-Szymańska and A. Boulton (eds.), Multiple Affordances of Language Corpora for Data-driven Learning. Amsterdam: John Benjamins, 109-128. SCoRE is available at www.score-corpus.org

[4] Frankenberg-Garcia, A. (2005). Pedagogical uses of monolingual and parallel concordances. ELT Journal, 59(3), 189-198.

[5] Frankenberg-Garcia, A. and Santos, D. (2003). Introducing COMPARA: The Portuguese-English parallel corpus. Corpora in Translator Education, 71-87. COMPARA is available at http://www.linguateca.pt/COMPARA/

[6] McEnery, A., and Xiao, R. (2007). Parallel and comparable corpora: What is happening. Incorporating Corpora. The Linguist and the Translator. Clevedon: Multilingual Matters, 18-31.

[7] No author. (n.d.) Linguee (website). Available at www.linguee.com

[8] No author. (n.d.) Tatoeba (website). Available at http://tatoeba.org/eng/

[9] No author. (n.d.) Reverso Context (website). Available at http://context.reverso.net/translation/

[10] St John, E. (2001). A case for using a parallel corpus and concordancer for beginners of a foreign language. Language Learning & Technology, 5(3), 185-203.

Please check the <u>Teaching Advanced Students course</u> at Pilgrims website.

Please check the <u>Practical Uses of Technology in the Classroom course</u> at Pilgrims website.

Please check the <u>Practical Uses of Mobile Technology in the Classroom course</u> at Pilgrims website.