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 <u>https://corpling4efl.wordpress.com</u>. E-mail: brown-ma@kifl.ac.jp

Menu

Introduction What are parallel corpora? Parallel corpora and language learning Conclusion References

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	(<u>www.score-corpus.org</u>) [3]
---------------------------------	-------------------------------------

	サンプリング なし 5 10	20 ソート 出現頁 左 キーワード 右 表示 KWIC センテンス
1	Every student was afraid of Mr. Smith .	どの学生もスミス先生を怖がりました。
2	assed out the information and told the students to read it .	彼らは資料を配布して, 生徒たちにそれを読むように言いました。
3	They thought he was an exchange ${\color{black}{\textbf{student}}}$ because they never saw him \cdots	彼らは今まで大学構内でまったく彼を見たことがなかったので,彼を交換留学生…
4	Pat number of teachers rated highly by $\ensuremath{\textbf{students}}$ in annual evaluations .	この学校には年次評価で学生から高く評価された教師たちが多数います。
5	This website can help students to schedule classes .	このウェブサイトは生徒たちが授業を予定表に組み込む手助けができます。
6	It is important to help students to succeed .	生徒たちが成功する手助けをすることが重要です。
7	ohn might go to Kyoto as an exchange student next year .	ジョンは来年交換留学生として京都に行くかもしれません。
8	If it was allowed , I 'd let $\ensuremath{\mbox{students}}$ use these software tools .	もしそれが許可されるなら,私は学生たちにこれらのソフトウェア・ツールを使…
9	The teacher found every one of his students quite bright .	先生は彼の生徒のだれもがとても賢いことに気づきました。
10	She is one of the smallest students in the class .	彼女はクラスの中で最も背の低い生徒のうちの1人です。

YV	VEBSCORE		Other Software About
Target	Language 🖲 English 🛇 Japanese	KWIC View: O Scrolling O Parallel	Keywords: Yes No
student	Search	Database: All Levels -	Show 🛨
Sort 1	CEN • Sort 2: 1L • Sort 3: 1R •	Sampled Hits: 10 -	
Cont I.			
HIT		Target Corpus	
1	Were there a lot of	students in the professor's office?	2
2	Were there a lot of	students who attended the lecture 1	last Monday?
□ 3	Several	students have gone into research an	nd development for a
4	; of competition between the two $\ensuremath{smartest}$	students in our grade.	
5	The school policy said that	students must wear uniforms.	
6	The teacher asked the	students to discuss what they found	d out about their
07	The assigned teachers gathered their	students and loaded them onto buses	3.
8	There are few new	students.	
9	school has given away t-shirts to several	students.	
□ 10	ichers practicing for the play amused the	students.	
ніт		Reference Corpus	
1	その 教授 の 研究 室 に は 大勢 の 学生 たち がい まし	たか?	
2	先週月 曜日 の その 講義 に 出席 し た 学生 は たくさん	, いましたか?	
□ 3	数人 の 学生 は ある 大手 の 製薬 会社 の 研究 開発 部	狎 へ 入社 しました 。	
4	我々の学年で成績トップの2人の生徒の間には	は常にある種の競争がありました。	
5	学校 の 方針 は 生徒 たち は 制服 を 着なけれ ば なら た	ぽいと書いてありました。	
6	先生 は 学生 たち に 彼ら の 研究 課題 について 発見 し	た こと を 議論 する よう 求め まし た 。	
7	配属 された 先生たちが 学生を集めて バスに乗せ	ました。	
8	ほとんど 新しい 生徒 は いませ ん 。		
9	学校 は 数人 の 生徒 に エシャツ を 配りました 。		
10	劇 の 練習 をし ている 教師 たちは 生徒 たちを 楽し	ませ まし た 。	

Image 2: WebSCoRE Interface (<u>www.antlabsolutions.com/webscore/</u>) [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



student			×E	English	🗸 ち Germ	an 🗸 🗸	Q	
			Θ					
	tudent mobilit	ty college stud	lent studer	nt body student	loans			
ranslations Student [nm.]	Schüler [nm.]	Studenten [n.]	Studentin In	f.] Schülerin [nf.]	Studierenden	studentischen	1	
Student [mir.]	Schuler [min.]	Studenten [n.]			Studierenden	studentischen	J	
studentische	studiert	Studenten- [adj.]	Studierende	e [nf.] Kursteilnehr	ner [n.] Teilneh	mer [n.] Studiu	m [n.]	
Studierender	[nm.] Teilnehr	merin [n.]						
A <mark>student</mark> O.I	D.ed on an uni	dentified hallucino	genic drug.	and the second s	arb an einer Über alluzinogenen Dr		nicht	
An NYU student was abducted last night.					Ein <mark>Student</mark> der New York University ist gestern Nacht entführt worden.			
A diligent student needs no teacher			Ein guter <mark>Schüler</mark> braucht keinen Lehrer, er ist selber eine					
- New <mark>student</mark> tryouts are next week.			- Die Aufnahmeprüfungen für die neuen <mark>Schüler</mark> sind nächste Woche.					
- New student tryouts are next week.			nächste Woch	e.	neuen <mark>Schüler</mark> Studenten von de			

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 <u>www.antlabsolutions.com/webscore/</u>

[2] Chujo, K., Kobayashi, Y., Mizumoto, A., & Oghigian, K. (2016). <u>Exploring the</u> <u>Effectiveness of Combined Web-based Corpus Tools for Beginner EFL DDL</u>. 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). <u>Pedagogical uses of monolingual and parallel</u> <u>concordances</u>. ELT Journal, 59(3), 189-198.

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

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

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

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

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

[10] St John, E. (2001). <u>A case for using a parallel corpus and concordancer for beginners of a foreign language</u>. 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.