

Translanguaging and data-driven learning: How corpora can help leverage learners' multilingual repertoires

Gaëtanelle Gilquin

UCLouvain, Belgium

<gaetanelle.gilquin@uclouvain.be>

Introduction

Translanguaging, i.e., the use of multiple languages to make and negotiate meaning, has been shown to be beneficial for language learning (see, e.g., García & Kleifgen, 2020). Although it is a fairly natural and spontaneous phenomenon in the lives of many multilingual speakers, its role is not well established in the language classroom, where the use of learners' mother tongues (L1s) or some additional languages (L2s) besides the target language (TL) is often frowned upon (by school authorities, teachers, and even students themselves). For this reason, there have been calls to "explore what 'teachable' pedagogic resources are available in flexible, concurrent approaches to learning and teaching languages bilingually" (Creese & Blackledge, 2010: 113).

One such possible resource is the corpus, an electronic database of naturally-occurring language that can be investigated by means of special tools and techniques to gain insights into the language(s) or language variety (varieties) represented in the corpus. What is particularly interesting about corpora from a pedagogical perspective is that they can be used by students to make their own discoveries about language, a pedagogical approach known as 'data-driven learning', or DDL for short (see Gilquin & Granger, 2022).

In what follows, I review some of the corpora that could be useful in translanguaging pedagogy and briefly show how they could help language learners leverage their own multilingual repertoires.

Corpora for translanguaging pedagogy

Translation corpora

The first type of corpus that could help develop students' translanguaging practices is the translation corpus, a corpus made up of texts in one language and their translation into one or several other languages. Provided they are aligned, translation corpora make it possible to search for a linguistic item and see its contextualized translations in the other language(s).

Translation corpora allow students to observe correspondences (or lack thereof) between the TL and some other language(s) they are familiar with. Teachers could thus encourage their students to look for occurrences of words known to be cognates or false friends in the other language(s) and to list their translations as they appear in the corpus.

The same sort of approach could be applied to phraseological expressions, including similes or proverbs, for example, so that students realize how close or different these expressions are in the multiple languages they speak. This could lead to discussions about cultural diversity.

Translation corpora can also be used to help students carry out translation tasks (cf. Laviosa, 2018), an activity that is said to promote translanguaging (García & Kleifgen, 2020).

L2 corpora

Corpora that represent non-native language varieties can be divided into three main types: learner corpora, which are made up of language produced by foreign/second language learners; lingua franca corpora, comprising language used as a lingua franca between speakers with different L1s; and corpora of institutionalized second-language varieties, i.e., varieties found in countries where the language has an official but non-native status (e.g., English in India).

All three types of L2 corpora can be used in translanguaging pedagogy, as they often include instances of other languages besides the TL (often the speakers'/writers' L1, but not only). In some of these corpora, such instances have been annotated in a special way, e.g., <foreign> in the Louvain International Database of Spoken English Interlanguage (LINDSEI; Gilquin et al., 2010) and <indig> in some components of the International Corpus of English (ICE; Greenbaum, 1996), so that they can automatically be retrieved from the corpora.

Investigating these types of corpora would show students how speakers/writers deploy features from their multilingual repertoires and for what purposes. In LINDSEI, for example, L1s/L2s are mostly used when the learner does not know the TL word, wants to refer to culturally-specific words, or produces discourse markers from the L1, e.g. 'enfin' in French or 'etto' in Japanese (Gilquin et al., 2010: 59-60). This could inspire students to use translanguaging for similar purposes or to think about other ways of approaching it. Seeing translanguaging in action may also ease the guilt that some students associate with it.

Writing process corpora

A relatively new type of corpus whose use can be advocated in translanguaging pedagogy too is the writing process corpus, which reproduces the processes through which a text is written, for example by means of keylogging and/or screencasting. This should make it possible to go beyond the product-oriented approach that is characteristic of translanguaging studies (Canagarajah, 2011).

A corpus such as the Process Corpus of English in Education (PROCEED; Gilquin, 2022) could be used by students to examine how writers rely on other languages while composing a text, be it as a temporary 'placeholder' (e.g., writing an L1 word when they cannot think of the TL equivalent) or when consulting an online bilingual dictionary, for example. Writing process corpora could also illustrate how secondary sources in multiple languages can be

drawn on to write a text in the TL (see S  ror & Gentil, 2020). Observing such processes could help students develop their own set of translanguage writing strategies.

Reference corpora

Finally, learners could be given access to reference corpora comprising model or ‘mentor’ texts (Garc  a & Kleifgen, 2020) that exemplify translanguageing. These could, for instance, be fictional works, academic articles, or political speeches produced by recognized authors or, more simply, conversations between fully competent bilingual speakers.

Not only can such corpora provide prime examples of translanguageing practices, but because of the status of their authors, they could help legitimize translanguageing in the eyes of learners.

Conclusion

Because corpora are potentially large databases representative of one or several languages/language varieties, they can be rich sources of information about translanguageing practices for students. Being computer-based, they allow for automation, which can facilitate the retrieval of relevant instances. They can be made up of spoken or written data, but also videos, which caters for a range of modalities. They can represent different genres and different languages/language varieties, which means that they can be used even in cases where teachers are not familiar with the repertoires used by their students.

By observing corpus data for themselves, students can become aware of the wide variety of translanguageing practices available and pick their favorites. This approach also comes with all the usual advantages of DDL (e.g., enhanced motivation, better cognitive processing, empowerment of students). If chosen carefully, corpora can present students with translanguageing they can relate to. This is particularly true if local corpora are used, that is, corpora collected in the students’ environment or possibly from the students themselves (e.g., outside the classroom, in contexts where they are perhaps less wary of mixing languages).

Although the focus here has been on language learners, it should be emphasized that trainee teachers too could benefit from corpora in their exploration of translanguageing practices. A corpus of teacher talk, for example, would be a good basis to illustrate how language teachers may combine the TL and the students’ L1 in the classroom and use each of them for different purposes (cf. Nicaise, 2021). Including this type of DDL in teacher training should help make future teachers more open to translanguageing (their own and that of their students).

While suggestions have been made about what corpora could be used in DDL-based translanguageing pedagogy and how, assessing the effectiveness of such approaches empirically and in different teaching contexts is crucial (cf. Canagarajah, 2011). If they turn out to be effective (in certain contexts), they could start to be adopted more widely by teachers who believe in the power of translanguageing.

References

- Canagarajah, S. (2011). Translanguaging in the classroom: Emerging issues for research and pedagogy, *Applied Linguistics Review*, 2, 1-27.
- Creese, A., & Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal*, 94(1), 103-115.
- García, O., & Kleifgen, J. A. (2020). Translanguaging and literacies, *Reading Research Quarterly*, 55(4), 553-571.
- Gilquin, G. (2022). The *Process Corpus of English in Education*: Going beyond the written text, *Research in Corpus Linguistics*, 10(1), 31-44. Available at <http://ricl.aelinco.es/first-view/174-Article%20Text-1066-1-10-20210407.pdf>.
- Gilquin, G., De Cock, S., & Granger, S. (2010). *The Louvain International Database of Spoken English Interlanguage. Handbook and CD-ROM*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Gilquin, G., & Granger, S. (2022). Using data-driven learning in language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics. Second edition* (pp. 430-442). London: Routledge.
- Greenbaum, S. (1996). *Comparing English worldwide: The International Corpus of English*. Oxford: Clarendon Press.
- Laviosa, S. (2018). Using corpora for contrastive analysis & translation-oriented ESP teaching. In S. Granger, M.-A. Lefer, & L. A. de S. Penha-Marion (Eds.), *Book of Abstracts. Using Corpora in Contrastive and Translation Studies Conference (5th edition). CECL Papers, 1*, 111-112. Louvain-la-Neuve: Centre for English Corpus Linguistics/Université catholique de Louvain.
- Nicaise, E. (2021). *Native and Non-Native Teacher Talk in the EFL Classroom: A Corpus-informed Study*. London: Routledge.
- Séror, J., & Gentil, G. (2020). Cross-linguistic pedagogy and biliteracy in a bilingual university: Students' stances, practices, and ideologies. *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 76(4), 356-374.