

Direct and indirect applications of learner and expert corpora to language teaching and learning: a top-down and bottom-up approach

This online workshop takes both bottom-up and top-down approaches to the study of language where both the Cambridge Learner Corpus (CLC) and expert language corpora are used. Both of these corpora are used in order to (1) better understand learner language use from different first languages, across levels, and in comparison to expert speakers and (2) develop teaching materials/develop teacher-training activities/create a DDL task informed by both corpus linguistics and language learning theories. This aim of this workshop is to create teaching and learning materials that could help learners of specific L1s acquire different and relevant items of language, and help those working in teacher training identify means of using corpora to help develop teachers. The corpora are housed in Sketch Engine which is the corpus analysis software used in this workshop.

The role of corpora for language learning is well established. Indirectly, within reference materials, the Longman Grammar of Spoken and Written English (Biber et al. 2012) and the Cambridge Grammar of English (Carter and McCarthy 2006) and within lexicography, the Collins learner dictionary (Collins 2016) and the Cambridge English Dictionary Online (2018), all exemplify corpus-informed resources used for language pedagogy. Further still, their use in corpus-based syllabi (Mishan 2005 p.44), corpus-based language textbooks like Touchstone (McCarthy et al. 2004-2006), language assessment (Kreyer 2015 p.14) and teacher training (Tsui 2006 p.57) show how wide reaching indirect applications of corpus linguistics to language teaching and learning have become. Directly, John's DDL (1990) or 'the learner as researcher' (Bernardini 2006 p.16) takes an inductive approach to language learning (Gavioli 2005 p.28), putting learners in contact with much more authentic language than traditional resources (Chambers 2010 p.345). The value of DDL, among other things, is its direct response to learner needs where DDL activities are often inspired by learner questions, that is, the things they want to know about their own language use and that of their language models (Hunston 2002 p.170). To this end, Meunier (2002 p.123) advocates for the use of learner corpora as well as expert speaker corpora when trying help learners learn a language. This use of both bottom-up and top-down approaches allows for contrastive interlanguage analysis (Granger 2015) which is invaluable to describing interlanguage, informing language learning directly, but also indirectly through informing materials such as those mentioned above.

The CLC is a 55-million-word corpus of students' written responses to Cambridge English exams. There are many different uses of this dataset. It gives us an insight into how learners use language and we can see how this changes over time where level can be seen as a proxy for time. An important value is also seen in the error tags that have been added to the texts to highlight the specific errors that the students have made in their writing. Analysing the data using these error tags gives valuable insight into the areas of language in which learners tend to perform well, and, where they tend to make mistakes. This information is used to inform ELT materials to make learners aware of common errors, so they can avoid making the same mistakes and focus on areas that require more attention. The metadata collected alongside the students' writing allows for the analysis of data specific to the students' first language (L1), the exam they took, which CEFR level they achieved, whether they passed or failed etc. There are over 173 L1s represented to varying degrees and the corpus continues to grow.

For this online workshop, we will begin with an overview of the aims, the corpora and the functionality of the software Sketch Engine. Next, groups will undertake research on the language of learners of different L1s and CEFR levels. Exploring the CLC, each group will identify different areas of

language of interest. The aim is to explore intuitions and discover how language is used by different learners. Next, in breakout rooms, the groups will have time to better understand the findings, digging deeper into the corpus and seeking explanations for their occurrence. Each group will then present their findings, applying them directly or indirectly to language teaching and learning. When the workshop is over, each participant will be emailed a copy of each presentation and task sheet produced in the workshop, as well as information about the corpora, the sample edition of the Cambridge Learner Corpus and how they can access it for research purposes.

For any questions/queries relating to this workshop, contact: ad3377@coventry.ac.uk.

References

- Bernardini, S. (2006) 'Corpora in the classroom: an overview and some reflections on future developments' in Sinclair, J. M., ed., *How to use corpora in language teaching*, Amsterdam: John Benjamins, 15-38.
- Biber, D., Johansson, S., Leech, G., Conrad, S. and E., F. (2012) *Longman grammar of spoken and written English*, Harlow: Pearson Education.
- Carter, R. and McCarthy, M. (2006) *Cambridge grammar of English: a comprehensive guide; spoken and written English grammar and usage*, Cambridge: Cambridge University Press.
- Chambers, A. (2010) 'What is data-driven learning?' in O'Keeffe, A. and McCarthy, M., eds., *The Routledge Handbook of Corpus Linguistics*, London: Routledge, 345-358.
- Collins (2016) COBUILD advanced British English learner's dictionary, available: [accessed 25 Mar 2016].
- Gavioli, L. (2005) *Exploring corpora for ESP learning*, Philadelphia: John Benjamins.
- Granger, S. (2015) 'Contrastive interlanguage analysis: A reappraisal', *International Journal of Learner Corpus Research*, 1(1), pp.7-24.
- Kreyer, R. (2015) 'The Marburg corpus of Intermediate learner English (MILE)' in Callies, M. and Götz, S., eds., *Learner corpora in language testing and assessment*, Amsterdam: John Benjamins, 13-34.
- McCarthy, M., McCarten, J. and Sandiford, H. (2004-2006) *Touchstone* (four levels), Cambridge: Cambridge University Press.
- Meunier, F. (2002) 'The pedagogical value of native and learner corpora in EFL grammar teaching' in Granger, S., Hung, J. & Petch-Tyson, S. eds., *Computer learner corpora, second language acquisition and foreign language teaching*, Philadelphia: John Benjamins, 119-142.
- Mishan, F. (2005) *Designing authenticity into language learning materials*, Bristol: Intellect.
- Tsui, A.B.M. (2006) 'What teachers have always wanted to know – and how corpora can help' in Sinclair, J. M., ed., *How to use corpora in language teaching*, Amsterdam: John Benjamins, 39-61.