

## **Invited Workshop: Learner Corpus Research: Merging Industry and Academic Interests**

Niall Curry and Olivia Goodman

Cambridge University Press

**Thursday, 11 September 2019, 14.30-18.00**

This workshop centres on the use of the Cambridge Learner Corpus (CLC) in order to better understand English language use from learners of different first languages and across levels of language competency. The corpus used in this workshop was created by Cambridge University Press and Cambridge Assessment English. All of the purpose-built corpora at Cambridge University Press are largely used to inform the development of their ELT materials, including general English course books, exam preparation materials and grammar references. The corpora are housed in Sketch Engine which is the corpus analysis software used in this workshop.

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 a 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 workshop, we will begin with an overview of the aims, the corpus 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, 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. 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.

**Niall Curry**, Cambridge University Press.

Niall is a Senior ELT Research Manager at Cambridge University Press and conducts research on language and language pedagogy to inform materials development with a focus on how we can use research from fields like corpus linguistics to better inform language learning. He is also completing his PhD at the University of Limerick, Ireland on corpus-based contrastive linguistics of academic writing in English, French and Spanish.

**Olivia Goodman**, Cambridge University Press.

Olivia is an ELT Research Manager at Cambridge University Press and supports author and editorial use of language research for content development; using corpora and bespoke language resources, her research ensures that materials are informed by real language insights. As a language teacher with a background in linguistics, Spanish and Mandarin Chinese from Nottingham Trent University, Olivia takes a research informed perspective for developing learning resources.