

## **Scrutinising the evidence: what can a data-driven approach contribute to our understanding of language use and development?**

**Geraldine Mark**

“Language looks rather different when you look at a lot of it at once.” (Sinclair, 1991)

This talk explores how a data-driven evidence-based approach can deepen our understanding of language use and development. With a particular focus on using learner corpora in language education, it aims to show how by systematically analysing large collections of authentic learner language, we can uncover patterns that are often invisible to intuition alone.

Learner corpora have been amassing for more than two decades (Granger et al. 2015). Research from this body of data has informed our understanding of morphology, lexis, syntax, and so on, across L1 and within L2 use. In more recent years, the connection between learner corpora and language development is being made and, increasingly, we are seeing emergence of data calibrated by level of competence (typically using the Common European Framework of Reference for Languages) rather than year of schooling. This has led to a knowledge base that describes typical usage and competencies of learners in relation to their language use at a given level of proficiency.

In this talk I'll draw on findings from The English Grammar Profile (EGP), a free educational online database, which provides a profile of over 1,200 corpus-based grammar competency statements about learner grammar use across the six CEFR levels. I'll describe the methodology used to build the EGP using the Cambridge Learner Corpus, discuss the key insights from the study and show how the investigation has enhanced our understanding of the developmental nature of grammar acquisition and use. I'll then look at further ways to explore the data taking a usage-based (UB) approach. UB studies have shown that language users are sensitive to the statistics of repeated patterns in language and that we figure out 'structural regularities' in language as we subconsciously tune into mappings of form and meaning (Ellis et al. 2016). Using this large scale proficiency-levelled data I will investigate if and how structural regularities develop in L2 English and how this might offer further insight into learner language development and inform language teaching.

### **References**

Ellis, N. C., Römer, U. & O'Donnell, M. B. (2016). *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Language Learning Monograph Series. Wiley-Blackwell.

Granger, S., Gilquin, G., & Meunier, F. (Eds.) 2015. *The Cambridge Handbook of Learner Corpus Research*. Cambridge University Press.

Sinclair, J. (1991) *Corpus, concordance and collocation*. Oxford University Press.