

Lexical Studies – Corpus Research Day 2015

Thursday 12th March

KEYNOTE 1: Michael McCarthy (The University of Nottingham)

Key concepts in spoken grammar

It is now 20 years since the first articles calling a shift of emphasis towards spoken grammars in language teaching began to appear in the applied linguistics literature. In this talk, I consider some of the main findings of corpus investigations into spoken grammar, focusing on several core interactive features of everyday conversation and examine these in terms of how they differ from typical written forms. Spoken grammars reveal a distinct set of features relating to the interpersonal stratum which underlies successful face-to-face communication. It is this stratum which, in the past, was often absent from language teaching materials and approaches to the teaching of speaking and the teaching of grammar.

Michael McCarthy is Emeritus Professor of Applied Linguistics at the University of Nottingham, Adjunct Professor of Applied Linguistics at the University of Limerick, and Visiting Professor at Newcastle University. He is co-director (with Ronald Carter) of the 5-million word CANCODE spoken English corpus project and the (co-)author and (co-)editor of 50 books and more than 100 academic articles, including the international best-seller adult course, Touchstone and its higher level, Viewpoint, as well as the Cambridge Grammar of English, English Grammar Today and several titles in the CUP English Vocabulary in Use series.



KEYNOTE 2: Paul Thompson (The University of Birmingham)

Corpus approaches to the language of interdisciplinary research articles

'Investigating interdisciplinary research discourse: the case of Global Environmental Change' (IDRD) is a cooperation between the Centre for Corpus Research, University of Birmingham and the international scientific publisher, Elsevier, and is funded by the Economic and Social Research Council (UK). The project focuses on the discourse of interdisciplinary research. We are employing a number of data-driven approaches to our corpus, which is made up of articles from eleven journals, six journals representing interdisciplinary research and five representing discipline-specific research. We are interested to find out, among other things, what can be found out by building up linguistic feature profiles and then seeing how the texts cluster. How do the texts cluster, and how can we account for the clustering? Do they cluster by research approach, by topic, or by discipline?

In this talk, I will discuss the results of a topic modelling exercise that we carried out on the data. Topic modelling is a machine learning technique that identifies 'topics' in a given corpus. We assume that a document consists of multiple topics with varying probability, and topic modelling estimates the distribution of topic probability in each document. From a topic model, we can then extract keywords of each topic, as well as the distribution of topics in each document, and we can cluster texts along different parameters. I will describe the procedures and then evaluate the results, both in terms of what they tell us about research discourse and in terms of how effective the technique is as a way of identifying useful keywords.

We also hypothesise that interdisciplinary research is typically 'outward-facing' in the sense that writers show an awareness of the needs of a readership that extends beyond their own discipline. To investigate this, we are looking at the uses of what have been called signalling nouns (Flowerdew and Forest 2015) or labelling nouns (Francis 1986); we do this first by determining their frequency and then analysing the ways that they are used, in one particular interdisciplinary journal, 'Global Environmental Change'. The analyses indicate that writers generally are concerned to provide textual signals for their readers and that one of the more frequently used such nouns in this journal is 'example'.

Paul Thompson is the Director of the Centre for Corpus Research at the University of Birmingham. His research interests are in academic and other specialised discourses, and in the exploitation of corpus resources and methodologies in learning about language. He is a Co-Editor of the Journal of English for Academic Purposes.



WORKSHOP 1: Laurence Anthony (Waseda University, Japan)

Analyzing Corpora with *AntConc*: From Basics to Best Practices

In this workshop, I will first provide a short history on corpus linguistics tool development highlighting the role that tools have played in the advancement of the field. Next, I will demonstrate some of the basic tools that corpus linguists use to analyze language data, including Key Word in Context (KWIC) concordancing, KWIC concordance plotting, cluster, n-gram, and collocate analyses, as well as word and keyword frequency generation. In real-world corpus linguistics studies, researchers often struggle to get corpus tools to provide the results they need. This may be because of mixed character encodings in the text files, the lack of part-of-speech tags in a corpus, or a host of other issues. To overcome these problems, the last part of the workshop will focus on some best practices in corpus linguistics research that will help researchers to quickly and easily analyze corpora in the way they want. For all the demonstrations and explanations of best practices, I will use the freeware, multiplatform corpus toolkit *AntConc* and several other freeware tools that I have developed.

Laurence Anthony is Professor of Educational Technology and Applied Linguistics and former Director of the Center for English Language Education (CELESE), Faculty of Science and Engineering, Waseda University, Japan. His main interests are in corpus linguistics tools development and English for Specific Purposes (ESP) program design and teaching methodologies. He is the developer of several corpus tools including AntConc, AntWordProfiler, AntMover, SarAnt, TagAnt, and VariAnt.



WORKSHOP/PAPER: Dawn Knight (Cardiff University)

DETAILS COMING SOON

***Dawn Knight** is a Senior Lecturer in Applied Linguistics at Cardiff University. Her current research interests lie predominantly in the areas of corpus linguistics, discourse analysis, multimodality and the socio-linguistic contexts of communication. The main contribution of her work has been to pioneer the development of a new research area in applied linguistics: multimodal corpus-based discourse analysis. This has included the introduction of a novel methodological approach to the analysis of the relationships between language and gesture-in-use based on large-scale real-life records of interaction (corpora).*

