

REPORTS FROM THE BRITISH ASSOCIATION OF APPLIED LINGUISTICS WITH CAMBRIDGE
UNIVERSITY PRESS SEMINARS 2021

RESEARCH IN PROGRESS

Corpora in applied linguistics: Broadening the agenda

Robbie Love*, Gavin Brookes and Niall Curry

Aston University, Birmingham, UK

*Corresponding author. Email: r.love@aston.ac.uk

1. Introduction

Due to the ongoing Covid-19 pandemic, the seminar, originally scheduled to take place at the Aston University campus, was hosted online. The aim of the seminar was to bring together researchers who are using corpora in applied linguistics (broadly defined); to facilitate discussion of the ways in which corpora are used in contemporary applied linguistics research, and how these may be used in future; and to explore the opportunities and challenges of engaging with stakeholders and achieving impact within applied corpus linguistic research.

There were over 200 registered participants from around the world, and the programme featured two plenary talks, 12 research papers and a panel discussion. The topics represented by the presentations in the programme included education, forensic linguistics, health (including Covid-19), oral histories, social media and tourism.

The seminar began with the first plenary, which was given by David Wright (Nottingham Trent University). Wright's talk outlined the role of corpus linguistics in some of the earliest and most high-profile cases in forensic linguistics and the growth of the field since the mid-1990s. He set out a roadmap for the continued application and expansion of corpus methods in forensic contexts that involved: (i) consolidating existing research and practice, (ii) seeking out and utilising already publicly available corpora and datasets and (iii) expanding the remit of 'forensic linguistics' to include new ways in which the relationship between language and law, crime, justice and evidence can be examined. To exemplify this latter point, he presented a case study in which he analysed the below-the-line comments on a *Mail Online* article reporting the early release of a terrorist from prison who went on to kill two people on London Bridge in 2019. He showed how established methods from corpus-assisted discourse studies can provide insights into public perceptions of the criminal justice system and how, in turn, such findings can broaden the scope of forensic linguistics.

Following the opening plenary were two research papers also on the theme of forensic linguistics. Firstly, Marton Petyko and Lucia Busso (Aston University) reported on research into latent topic changes in the *Operation Heron* abusive letter series. Using a combination of Structural Topic Modelling and qualitative corpus analysis, involving the extraction of topic-related words and n-grams, they examined a corpus comprising 9,165 words of text from 50 letters and 49 envelopes, sent between January 2007 and April 2009. Petyko and Busso demonstrated how their novel approach can be used to identify topic-specific arguments and elicit hidden themes in forensic texts. Then, Mark McGlashan (Birmingham City University) presented an analysis of children's online disclosures of abuse to *Childline*, with the aim of identifying and interpreting the relationships between different types of abuse and the linguistic choices within the disclosures. Investigating a corpus of 3,242 posts disclosing domestic abuse, McGlashan demonstrated how disclosures are often complicated and highly context-dependent, and discussed how this work could be of direct use to relevant practitioners in better supporting children to make such disclosures.

© The Author(s), 2021. Published by Cambridge University Press

In the next panel were papers by Anna Čermáková (University of Cambridge) and Chris Fitzgerald (University of Limerick). Čermáková presented an investigation into vague language in classroom talk. Her talk centred on the affordances of vague language for facilitating socialised learning. Through the use of the Dialogue and Argumentation for Cultural Literacy Learning in Schools (DIALLS) corpus and its Key Stage 2 (KS2) subcorpus, she retrieved a number of key examples of vague language, such as the quotative *like*, and signalled the value of such vague language to education by highlighting the role it plays in students' knowledge construction processes. Subsequently, moving away from educational applications of corpus linguistics, Fitzgerald presented an investigation into epistemic modality in Irish historical narratives. His paper set in relief a number of key synergies that exist between historical linguistics, corpus linguistics and oral histories. In discussing the complex process of constructing the Corpus of Irish Historical Narratives, he addressed issues in oral transcription, challenges in working with historical documents, issues of accuracy in the documentation of past events, as well as variability in the role of epistemic modality therein.

The final panel of Day 1 featured two papers concerning the analysis of data from Twitter. Firstly, Sten Hansson and Ruth Page (University of Birmingham) presented their ongoing work utilising a corpus-assisted approach to examine blame avoidance in tweets from UK government departments and political figures. In particular, they assembled a corpus representing tweets relating to the UK's departure from the European Union and governmental policy responses to the Covid-19 pandemic. The data were manually coded for a wide range of blame avoidance strategies, identified based on an existing framework, which indicated that the government legitimated its actions using a series of legitimising appeals. The appeals could be effectively identified using keywords, with some appeals visible through searches of single keywords, while others could only be identified through searches of larger phraseological patterns. In the second talk of this panel, Saira Fitzgerald (Lancaster University) presented an analysis of Twitter discourses surrounding the International Baccalaureate in the context of the Covid-19 pandemic. This talk reported on the analysis of 4,278 tweets featuring the hashtag, *#ibscandal*, posted over a two-month period in 2020. Focusing on nomination and predication strategies, this analysis uncovered shifts in discourses that are intertextually linked to events in the wider world that, Fitzgerald argued, provide rare insight into how students are impacted by the global education industry.

Day 2 of the programme commenced with the second plenary talk, presented by Paula Buttery (University of Cambridge). Buttery discussed the uses of corpora in education technology and delivered an interdisciplinary perspective on the development of corpus-informed language correction technologies. Focusing on technologies such as *Write & Improve* and *Speak & Improve*, the plenary discussed how computational linguistics, conversation analysis, and corpus linguistics can offer a rigorous and multidirectional approach to addressing extant challenges in language education and the needs of the language education industry. Focusing on the power of computational models for predicting errors, corpus linguistic research for contextualising and qualifying such errors, and conversation analysis for informing discourse patterns, Buttery exemplified how these three disparate fields can work in synchrony. Pointing forward, the plenary signalled the need for interdisciplinary thinking, critical perspectives and industry engagement for the development of further language education technologies.

Following the plenary, Peter Crosthwaite (University of Queensland) and Kamonchanok Sanmuang (Kasetsart University) presented papers that continued with the theme of education. Crosthwaite discussed the current state of the art with regards to corpora and data-driven learning (DDL) for young learners, surveying the benefits of DDL as well as the barriers to implementation in the primary and secondary classroom. He argued that, while young learners are used to living in a digitally-connected world, this does not necessarily translate into the proficiencies necessary for successful engagement with corpus data. Following this, Sanmuang discussed the pedagogical applications of research into specialised lexical frames for undergraduate public health students in Thailand. Based on the analysis of four-word n-grams in a 1.3-million-word corpus of academic articles from the PublicHealth

Corpus, Sanmuang developed targeted linguistic instruction and demonstrated the effectiveness of such an approach for improving proficiency in specialised academic language.

The next panel focused on the theme of health communication. First to present was Luke Collins (Lancaster University), who discussed the application of corpus methods to the investigation of voice-hearing. Collins used corpus methods to examine semi-structured interviews with 67 voice hearers, focusing on aspects of the participants' responses that are relevant to a better understanding of voice-hearing, including continuity between the experiences of clinical and non-clinical voice-hearers; personification of voices; and impoliteness and rapport building. The paper concluded with reflections on the process of collaborating within interdisciplinary teams and a discussion of some of the implications of the project findings for clinical interventions designed to support voice-hearers to live well with their voices. In the second talk of this panel, James Balfour (Lancaster University) presented an analysis of schizophrenia as a metaphor in the British Press. Analysing a corpus of UK newspaper articles published between 2000 and 2015, Balfour examined the collocates of the noun SCHIZOPHRENIA and found that the term was used metaphorically in relation to a range of topics, including finance, fashion and sport. Balfour also adopted a diachronic perspective, finding that metaphorical meanings of schizophrenia have expanded over time and in conjunction with changing public attitudes around the condition. Balfour concluded by reflecting on the complex relationship between metaphor, etymology, and historical and contemporary understandings of mental illness.

The final presentation panel of the programme featured papers from Karoline Irschara (University of Innsbruck) and Valeria Franceschi (University of Verona). Irschara presented an investigation of gender in radiology reports. Offering a non-English language perspective on a less-studied medical genre, this paper discussed language patterns and collocations, with a view to unpacking how male and female patients are discussed similarly and differently in this context. Among the findings was a focus on administrative processes and issues of consent among female patients, and descriptive and past references such as HYDRATED OR REACTED among male patients. Then, Franceschi discussed the creation of a corpus of destination video blogs. Her paper carried forward the focus on specialised discourses in less-studied genres and offered a further, multimodal dimension. In discussing the video blog genre, the paper addressed the challenges evident in identifying the parameters of such a recent and variable genre, and highlighted specific linguistic features that play critical roles in organising the discourse and modalities in destination video blogs.

2. Outcomes and implications for applied linguistics

The programme ended with a panel discussion, chaired by Robbie Love, which featured several of the speakers from both days of the seminar. Several topics were discussed, including the challenges and opportunities of working and publishing with colleagues from other disciplines; ethical issues with regards to collecting and analysing sensitive data; and new directions and future developments in applied corpus linguistics. One of the main outcomes of the seminar was a shared understanding that the opportunities for corpus methodologies to contribute valuably to contexts outside the realm of linguistics are only growing.

The aim of the seminar was to bring together a diverse group of researchers to share perspectives on the applications of corpus linguistics in a range of disparate contexts. In doing so, we afforded the opportunity for participants to consider some current methodological innovations as well as broad challenges that unite scholars in the field, regardless of specific inter-disciplinary application.