**Corpora, Discourse and Society**

BAAL Corpus Linguistics SIG Annual Workshop

13th November 2019

Lancaster University

**Programme**

|  |  |
| --- | --- |
| **09.00 – 09.30** | **Registration and welcome** |
| 09.30 – 10.30 | **Plenary 1: Lesley Jeffries**Beyond the Concordance Lines: making corpus methods work for discourse analysis |
| **10.30 – 10.50** | **Coffee break** |
| 10.50 – 11.20 | Tatyana Karpenko-Seccombe*Language of conflict: cross-linguistic corpus-assisted comparative discourse study* |
| 11.20 – 11.50 | Isobelle Clarke and Mark McGlashan*Online discourses of toxic masculinity* |
| 11.50 – 12.20 | Mel Evans*Literary Lifespans: corpus perspectives on linguistic repertoires in early modern drama* |
| **12.20 – 13.30** | **Lunch and posters** |
| 13.30 – 14.30 | **Plenary 2: Tony McEnery***The UK, Europe and the path to Brexit - the long view: Europe in two centuries of British newspapers.* |
| 14.30 – 15.00 | Duygu Candarli, Robbie Love and Alice Deignan*Rhetorical variation in teachers’ lesson presentations across Key Stage 2 and Key Stage 3* |
| 15.00 – 15.30 | Michael Handford*‘The Russians came along with a bucketful of Roubles’: National identity markers in narratives in business meetings* |
| **15.30 – 16.00** | **Coffee break** |
| 16.00 – 16.30 | Lexi Webster*“Something from Nothing”: Towards a corpus-informed cultural political economy analysis* |
| 16.30 – 17.00 | Susan Whitbread*Corpus-assisted discourse analysis of opposition to GMOs in food and agriculture* |
| 17.00 – 17.30 | Caterina Guardamagna, Victorina González-Díaz and Ursula Kania*“We talk of birthright”: A corpus assisted discourse analysis of Early Modern representations of ‘citizenship’ (1550-1700)* |
| **17.30 – 17.45** | **Wrap up and close** |

**Posters**

|  |  |
| --- | --- |
| **Presenter** | **Title** |
| Ondřej Molnár | Textual Rhetoric of News Stories |
| Charlotte-Rose Kennedy | The representations of the People’s Vote March in *The Express* and *The Independent* |
| Luke Collins | The discourse of PrEP in the UK Press (2014-2019) |
| Gillian Smith | Teacher directives and pupils responses in SEN classrooms: insights from corpus methods |
| Augmentative and alternative communication (AAC) in the Special Educational Needs classroom: a corpus analysis of small group literacy teaching. |
| Lorrae Fox | Examiner and candidate collocation use in the Trinity Lancaster Corpus |

Note: This list is provisional and more posters may be added!

**Plenary abstracts**

**Beyond the Concordance Lines: making corpus methods work for discourse analysis**

Lesley Jeffries

*University of Huddersfield*

Though corpus methods have long been just one part of my repertoire of linguistic methodologies, in recent years I have focused on how to take the advances in corpus linguistics and make them work for me. As a linguist interested in textual meaning, I have become fascinated by how individual lexical items in particular can be buffeted by the relatively local events of socio-political life – in both time and space – in order to observe what I have sometimes called ‘emergent meaning’.

One of the consequences of the ubiquity of our modern (traditional and social) media is that words can take on specific and sometimes very powerful roles in the public debates that affect the ways in which we are governed and how our society works. Identifying and tracing the way in which such short-term and local semantic change happens – and the potential effects it may have – is my general research objective in this work. My work with colleagues (Brian Walker, Dan McIntyre and Matt Evans) has taken the investigation of what we are calling ‘sociopolitical keywords’ as the testing ground for how to bridge the gap between the techniques of data manipulation provided by corpus software and the need to scrutinize the co-text in detail in order to see how the word is used.

Whilst the research strengths of corpus methods (replicability, rigour, transparency) are evident, though possibly not always as objective as claimed, the same cannot be said of discourse analysis, whose aim is so broad that such expectations buckle under the weight of contextual information. In this talk, therefore, I’d like to explore the challenges in corpus-based work once the concordance lines have been produced. Drawing on my theoretical work in critical stylistics, I will propose one way in which these challenges can be met.

**The UK, Europe and the path to Brexit - the long view: Europe in two centuries of British newspapers**

Tony McEnery

*Lancaster University*

When the UK voted to exit the European Union in 2016 many tried to understand the event. Opinion polls, the analysis of voting patterns, studies of the press in the run up to the vote, the role of social media and machine learning all came to prominence as people and organizations tried to understand what was an unexpected outcome for many.

In the research presented in this paper, undertaken with Helen Baker and Vaclav Brezina, we approached this question with a different mindset. Much of the work we have done in the past has been focused on exploring the development of attitudes and ideas, as reflected in language, over long stretches of time. We believe that in taking this approach we can start to understand long term trends that give rise to specific, notable, events.

Brexit was not the result of one or two months of campaigning. Brexit needs to be understood in the long term. Only by doing so can we begin to answer questions such as ‘How did the UK vote for Brexit when no major political party in the UK has ever campaigned to leave the EU?’, ‘How and why did attitudes to the EU change in the UK between it voting to stay in the EU in 1975 and voting to leave in 2016?’ and ‘What were the key events over time that led to Brexit?’.

In this talk, we will approach questions such as these in two ways. After a brief review of the history of British engagement with what is now called the EU, we will start by summarising what researchers in academic disciplines such as History and Politics say have been the major arguments for and against staying in the EU over time. We will follow that up by a study of one newspaper, from the 1960s to the present – The Times. We have access to all of the machine readable copies of the newspaper – over three billion words were published by it in this period alone. Using techniques pioneered by linguists to look at and start to comprehend data on this scale using computers, we will explore the totality of this data to look at how attitudes to the EU shifted over time and link those to historical events. By doing this we will start to show, for the first time, on a very large scale how British attitudes to the EU developed over time, as viewed through the prism of one very influential newspaper. In doing so we will also reflect on the claims made by the smaller scale, qualitative research, undertaken by Historians and Political Science researchers while gaining a deeper understanding of the drivers of Brexit.

**Panel session abstracts**

**Language of conflict: cross-linguistic corpus-assisted comparative discourse study**

Tatyana Karpenko-Seccombe

*University of Huddersfield*

2014 was undeniably a year of dramatic events in the history of Ukraine, above all the Russian annexation of Crimea and the formation of two self-proclaimed pro-Russian Republics which led to full-scale armed conflict between Ukrainian government troops and pro-Russian forces, followed by a failed attempt to negotiate a truce and a ceasefire.

Against the backdrop of these events, this study focuses on the parliamentary debates of the two countries in conflict. By comparing corpora of pre- and post-conflict debates, the study explores the way the socio-political and historical context in which the word use is set triggers shifts in word meanings. The paper presents a cross-linguistic, corpus-assisted discourse analysis of parliamentary debates, where keyword and collocation analysis is used to examine the shifts in connotative, affective and contextual word meanings and the effects these shifts have on parliamentary discourse in both countries. Part of a larger discourse analysis project, the paper, however, concentrates on three words only: *separatist, truce* and *brother,*  which are indicative of such shifts.

Comparative analysis of the use of Ukrainian and Russian translation equivalents *сепарат\*/сепарат\*/separat\** has shown movement of word meaning from neutral to negatively-coloured reflecting the ideologies and attitudes of the speakers. Collocates of *перемир/перемириe/truce* have developed negative connotative meanings associated with the failed truce of 2014, demonstrating the rise of the negative associative contextual meanings in neutral words. Analysis of *брат\*/брат\*/brother* shows a distinct shift from neutral literal meaning to negative figurative meanings in parliamentary discourse. The corpus data show the way the affective and connotative meanings have begun to diverge in the discourses of both parliaments.

In conclusion, such cross-linguistic corpus-assisted approach may become a productive tool for studies in the language of conflict within the framework of conflict resolution.

**References**

Jeffries, D., O'Driscoll, J., Evans, M. (2018). ‘Language in conflict: Linguistics in mediation’. In: D. McIntyre and H. Price (eds*), Applying linguistics: Language and the impact*

*agenda*. London and New York: Routledge.

McEnery, A.M. and Baker, P. (eds) (2015). *Corpora and Discourse Studies: Integrating*

 *Discourse and Corpora*. Basingstoke: Palgrave Macmillan.

**Online discourses of toxic masculinity**

Isobelle Clarke and Mark McGlashan

*University of Birmingham / Birmingham City University*

*Toxic masculinity* means different things for different people in different contexts, and with the influence of the media it has come to be used as a catch-all term for harmful behaviours and attitudes of men informed by (hegemonic) masculinity, especially those related to power and inter-/intra-group conflict on the basis of gender (e.g. misogyny) and sexuality (e.g. homophobia). Whilst some research has begun to acknowledge the term’s utility in delineating culturally accepted and social destructive aspects of hegemonic masculinity (Kuller, 2005), there has not yet been an account of what *toxic masculinity* is and what it means. This paper aims to provide an initial linguistic account of those aspects and behaviours that *toxic masculinity* is (and *can be*) used to refer to by investigating *how* it is used.

Consistent collocates

Figure 1

A specialised ~900,000 word corpus was built that contains two subcorpora of posts from the social media websites Twitter and Reddit including the term *toxic masculinity*. Analysis began by identifying collocates of the terms *toxic* and *masculinity* (5L-R span, MI >= 3, min. frequency = 20) in each subcorpus. An intersection of these lists was then performed to determine a set of statistically significant and *consistent collocates* (Figure 1) across both subcorpora. *Consistent collocates* were used as a starting point for analysis and explores topics relating to:

* gender (*men, gender, femininity, patriarchy*)
* feminism (*feminists,* feminism, feminist)
* meanings and concepts (*term, think, thing, example, mean, means, concept, idea, something, nothing*)
* existence (*exist, exists*)
* speech (*saying, say, talking, called*)
* problems (*problem, bad*)

Analysis of these topics consistently associated with *toxic masculinity* enable us to offer a socially constructed definition of *toxic masculinity*.

**References**

Kupers, T. A. (2005). Toxic masculinity as a barrier to mental health treatment in prison*.*

 *Journal of Clinical Psychology*, 61, 713–724.

**Literary Lifespans: corpus perspectives on linguistic repertoires in early modern drama**

Mel Evans

*University of Leicester*

This paper considers how corpus-based analyses of dramatic language can contribute to sociolinguistic research; specifically, our understanding of language change across the lifespan. Dramatic language is an established resource for historical linguistic research (e.g. Culpeper and Kytö 2010); relatedly, in literary stylometry (e.g. Craig and Hirsch 2016) corpora of early modern drama have been the focus of scholarly debate and innovation surrounding authorship and style. The notion of literary ‘authorial style’ is a popular, as well as forensically useful, concept, as is the idea that literary style changes over time. These concepts have parallels in sociolinguistic theories of non-literary language, in speaker repertoires and lifespan variation and change.

My paper thus explores ‘literary linguistic lifespans’, considering the repertoires of three Restoration (c. 1660-1700) dramatists: Aphra Behn, John Dryden and Thomas D’Urfey. I focus on their use of interjections: a word class important in the depiction of orality and interpersonal relationships (e.g. Taavitsainen 1997; Culpeper and Kytö 2010), identifying the socio-pragmatic attributes of interjections’ functions and forms across each authorial corpus. The results suggest that authorial preferences are distinctive: repertoires change in size, and show indiosyncratic diversity in their functional scope, working with but also against macro-level trends. In bringing the literary to bear on the sociolinguistic, corpus methods provide a valuable interdisciplinary bridge with significant potential.

**References**

Craig, D. H., and Brett Greatley-Hirsch. 2017. *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge: Cambridge University Press.

Culpeper, Jonathan, and Merja Kytö. 2010. *Early Modern English Dialogues: Spoken Interaction as Writing*. Cambridge: Cambridge University Press.

Taavitsainen, Irma. 1995. ‘Interjections in Early Modern English: From Imitation of Spoken to Conventions of Written Language’. *Historical Pragmatics: Pragmatic Developments in the History of English* (ed. by Andreas Jucker). Amsterdam: John Benjamins, 439 -66.

**Rhetorical variation in teachers’ lesson presentations across Key Stage 2 and Key Stage 3**

Duygu Candarli, Robbie Love and Alice Deignan

*University of Leeds*

Lesson presentations are a common resource in primary and secondary schools in England. They form a hybrid register that can include lesson objectives, subject content, questions, and assessment information. An earlier corpus study by our team (Candarli et al., 2019) found that there is a marked change in the academic language used in texts intended for students between KS2 and KS3. This is likely to pose a challenge to many students starting KS3, which, ideally, a lesson presentation that follows a known and understood structure could help to mitigate.

This study aimed to investigate rhetorical similarity and variation in teachers’ lesson presentations across KS2 and KS3. We used ProtAnt (Anthony & Baker, 2015) to select prototypical texts in corpora of KS2 and KS3 lesson presentations. In those texts we identified moves and steps (Swales, 1990) and tagged lexico-grammatical features (Nini, 2014) to reveal which lexico-grammatical features convey the rhetorical functions of moves.

KS3 lesson presentations are more complex than those in KS2 and include a greater variety of communicative functions, some of which are not present, or are rare, in KS2. This could mean that their purpose and meaning is unclear for some students moving into KS3. The potential of the lesson presentation to frame and support comprehension of the new linguistic and subject material of KS3 may be lost. The implications of this for the school transition will be discussed.

This study is part of the ESRC-funded project The linguistic challenges of the transition from primary to secondary school (grant ref: ES/R006687/1; project website: https://linguistictransition.leeds.ac.uk/).

**References**

Anthony, L., & Baker, P. (2015). ProtAnt: A tool for analysing the prototypicality of texts. International Journal of Corpus Linguistics, 20(3), 273-292. DOI: 10.1075/ijcl.20.3.01ant

Candarli, D., Love, R., & Deignan, A. (2019). Using corpora to explore the language challenges of the transition from primary to secondary school. Corpus Linguistics 2019. Cardiff University, UK. July 2019.

Nini, A. (2014). Multidimensional Analysis Tagger 1.2 - Manual. Retrieved from: http://sites.google.com/site/multidimensionaltagger.

Swales, J. (1990). Genre analysis: English in academic and research settings. Cambridge: Cambridge University Press.

**‘The Russians came along with a bucketful of Roubles’: National identity markers in narratives in business meetings**

Michael Handford

*Cardiff University*

This paper explores how interlocutors may, or may not, make national identity relevant in business meetings discourse. As such it extends Piller’s (2011) exhortation to explore why, how and when people make national identity relevant in business, to the analysis of naturally occurring narratives in meetings. Narratives are defined here as ‘small stories’ (Bamberg and Georgakopoulou, 2008), that is, narratives that may not follow a ‘canonical’ structure, and which cannot be assumed to be straightforward, unmitigated representations of past experiences (Bamberg and Georgakopoulou, 2008); instead, narratives are seen as ‘aspects of situated language use, employed by speakers/narrators to position a display of situated, contextualized identities’ (ibid: 377).

The study examines the million-word CANBEC corpus (Handford, 2010) of business meetings for instances of appropriate narratives. Methodologically this involves first pinpointing instances of ‘explicit national identity markers’, for instance ‘Pascal’s very French’, where a person’s or a group’s identity is framed in terms of nationality. As such, it examines how nationality is made relevant in discourse, rather than assuming a priori that it is relevant, and moves beyond essentialist conceptualisations of national identity which may be prevalent in much corpus linguistic and sociolinguistic research. Next, extended concordance lines are examined for potential narratives. This talk examines a selection of relevant narratives pinpointed through the concordance lines, and discusses what they reveal about the situated identities and ideologies of the interlocutors, as well as the relationship between national stereotypes and narratives.

**References**

Bamberg, M., and Georgakopoulou, A. 2008. Small stories as a new perspective in narrative and identity analysis. Text and Talk, 28(3): 377-396.

Handford, M. 2010. The Language of Business Meetings. Cambridge: Cambridge University Press.

Piller, I. 2011. Intercultural Communication: A Critical Introduction. Edinburgh: Edinburgh University Press.

**“Something from Nothing”: Towards a corpus-informed cultural political economy analysis**

Lexi Webster

*Manchester Metropolitan University / Lancaster University*

This paper operationalises a novel approach to critical discourse analysis, combining corpus tools and techniques with approaches to political economy to identify barriers to and opportunities for social reform.

In this paper, I refer to two approaches to political economy that share a critical imperative for social change. Cultural political economy (CPE) considers how imaginaries, and discourses therein, are reproduced (e.g. Jessop & Oosterlynck, 2008). If sufficiently retained and reproduced, selected discourses become ‘[integrated] into patterns of structured coherence’ (Jessop, 2010: 341), and illuminate how social imaginaries are both organised and governed (Jessop, 2004). With this information, CPE aims to critique harmful social formations. Political economy analysis, then, seeks reform opportunities by identifying *inconsistencies* in the ‘prevailing political and economic processes of society’ (McLoughlin, 2014: 5). Hence, both approaches attempt to understand the hegemonic structures of an imaginary and identify opportunities for positive change. However, neither approach fully operationalises a method for identifying hegemonic processes and/or discourses.

Using corpus linguistic techniques, this novel approach takes a holistic view of hegemonisation and analyses discursive dominance in three ways. That is, I identify: (1) characteristic discursive structures per keyness scores; (2) *horizontally* dominant discourses frequently reproduced by many actors, and; (3) *vertically* dominant structures granted authoritative status via political or economic means. Inconsistencies between the three levels of hegemonisation constitute barriers to and opportunities for reform.

**References**

Jessop, B. 2004. Critical semiotic analysis and cultural political economy. *Critical Discourse Studies*, 1(2). 159–174.

Jessop, B. 2010. Cultural political economy and critical policy studies. *Critical Policy Studies*, 3(3–4). 336–356.

Jessop, B. & Oosterlynck, S. 2008. Cultural political economy: On making the cultural turn without falling into soft economic sociology. *Geoforum*, 39(3). 1155–1169.

McLoughlin, C. 2014. Political economy analysis: Topic guide. *Governance and Social Development Resource Centre,* University of Birmingham, UK.

**Corpus-assisted discourse analysis of opposition to GMOs in food and agriculture**

Susan Whitbread

*Monash University, Australia*

The use of genetically modified organisms (GMOs) in food and agriculture is one of many issues that people argue about and take sides on publicly. Such oppositional discourse is generally analysed within a broader framework of politeness/impoliteness scholarship (Graham, 2017; Jeffries 2010).

Increasingly, web 2.0 (the interactive web) provides new ways of interacting and arguing online. Although analysis of GM discourse is not a new topic (Cook, 2004; Doolin, 2007), the relatively recent arrival of sophisticated gene editing tools such as CRISPR/Cas9 and gene drives—together with renewed government policy and regulatory interest, particularly within the Australian context—provides a timely opportunity to take a fresh look at the linguistic resources and strategies people bring to online discourse and debate.

To ensure an objective approach to data selection and analysis I have undertaken a corpus-assisted discourse analysis (Mautner, 2015) of online submissions to a number of Australian government reviews and inquiries into gene technologies, which have been taking place since early 2016. Submissions were downloaded and compiled into a purpose-built corpus in Sketch Engine (Kilgarriff et al, 2014) and analysed against a global English web reference corpus (enTenTen15) in Sketch Engine. Frequency, concordance, keyword and collocation analyses were undertaken and selected texts were examined to ascertain their argumentative structures and resources.

Early results suggest that although an individual submission may not be overtly argumentative in nature, there is linguistic evidence to link back to well-established public controversies around GM.

**References**

Cook, G. (2004). Genetically modified language (Kindle edition). Abingdon: Routledge

Doolin, B. (2007). Biotechnology discourse. Discourse Studies 9(1) pp.5-8.

Graham, S.L. (2017) Politeness and impoliteness. In C. R. Hoffmann and W. Bublitz (eds.) Pragmatics of Social Media, pp. 459–491. Berlin/Boston: De Gruyter Mouton

Jeffries, L. (2010). Opposition in discourse. London, UK: Bloomsbury Academic

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. and Suchomel, V. (2014) The Sketch Engine: Ten Years On. Lexicography, 1. pp7-36.

Mautner, G. (2015) Checks and balances: how corpus linguistics can contribute to CDA. In Wodak, R. & Meyer M. (eds) Methods of critical discourse studies (3rd edition). Los Angeles: Sage.

**“We talk of birthright”: A corpus assisted discourse analysis of Early Modern representations of ‘citizenship’ (1550-1700)**

Caterina Guardamagna, Victorina González-Díaz and Ursula Kania

*University of Liverpool*

This study, developed in dialogue with the ERC-funded project *Travel, Transculturality and Identity* in England 1550-1700 (TIDE), aims at providing new insights on issues of identity and belonging in Early Modern England. By taking a corpus-based discourse analytical approach (Baker 2006), this study offers a complementary perspective to TIDE, whereby relevant social controversies are explored through the lens of “cultural keywords” (Williams 1976) investigated through the close reading of a small number of significant texts. Using TIDE keywords as entry-points, a set of relevant keywords was arrived at by consulting the OED and contemporary dictionaries (LEME) on the one hand and observing statistically-salient related terms in EEBO on the other. This paper – primarily based on EBBO - focuses on a contrastive analysis of terms related to the semantic field of citizenship: subject, citizen, burgess, freeman.

Initial results show a complex picture. Whereas all free-born individuals are subjects whose political status is defined by an oath of allegiance to the king, only a sub-set of subjects are citizens, enjoying additional privileges. Both citizens and subjects are strongly characterized by a moral dimension founded on the values of obedience and deference. A spiritual dimension is strongly associated with citizen (‘citizen of God’) and to an extent with freeman (‘the Lord’s freeman’) – but is absent from burgess. Social status defined by wealth is an important feature among citizens, whereas being poor is concern for those subjects who may be recipients of alms. Wealth does not appear as a defining feature of burgesses.

This research contributes to a larger study including three main social categories: an in-group (e.g., citizens, settlers), an out-group (e.g. foreigners, strangers, aliens) and an in-between group of ‘movers’ (e.g. travellers, vagrants) shaping the complex picture of Early Modern society. This study shows that also the in-group is a complex mosaic of social identities (see Borot 2016).

**References**

Baker, P. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.

Borot, L. (2016) ‘Subject and Citizen: The Ambiguities of the Political Self in Early Modern England.’ *Revue Française de Civilisation Britannique*, XXI-1.

\*\*\*

Citation in the title: Henry Ireton, speech held on 29 October 1647.