CAD 2018

Organising Committee

- Paul Baker (chair)
- Carmen Dayrell
- Tony McEnery
- Andrew Hardie
- Rachael McCarthy (administrator)

Programme Committee

- Paul Baker (Lancaster University, UK)
- Carmen Dayrell (Lancaster University, UK)
- Alison Duguid (University of Siena, Italy)
- Costas Gabrielatos (Edge Hill University, UK)
- Kevin Harvey (The University of Nottingham, UK)
- Tony McEnery (Lancaster University, UK)
- Alan Partington (University of Bologna, Italy)
- Gerlinde Mautner (Vienna University, Austria)
- Sylvia Jaworska (University of Reading, UK)
- Sally Hunt (Rhodes University, South Africa)
- Laura Paterson (The Open University, UK)
- Vaclav Brezina (Lancaster University, UK)

PLENARY TALKS

Making a patchwork quilt or Frankenstein's monster?
Interdisciplinarity and collaboration in CADS

Amanda Potts
Cardiff University
pottsa@cardiff.ac.uk

Interdisciplinarity is all the rage; in conference calls, grant applications, and job interviews, scholars are increasingly asked to detail the ways in which their work crosses disciplinary boundaries. This seems to support a tacit belief that interdisciplinary work is incontrovertibly 'better'. But is this always the case? Collaboration is an art form, and crossing disciplines can mean losing language and focus, as well as generating misunderstandings around disciplinary cultural 'norms'. While collaborating with teams, individual scholars may feel that their individual voices are muffled or their areas of interest are not adequately (or fairly) represented. When work is complete, it can be difficult to place for grants, journals, and conferences. CADS is, in itself, an interdisciplinary field and finds itself in many intersections with other disciplines. This gives us a unique vantage point from which to consider the potential strengths and pitfalls of interdisciplinary collaboration.

In this talk, I will discuss the ways that I and others have integrated interdisciplinary voices into CAD scholarship, including: guidance on sampling; interpretation of results; and triangulation of findings using alternative frameworks. Examples will be presented from my latest work, which has been in collaboration with social media content creators and with jurists. I will talk about the strengths and limitations (as I see them) of taking interdisciplinary approaches to these data and the lessons that I am still learning about working in a team without losing one's hard-fought disciplinary identity.

Amanda Potts is a Lecturer in the Centre for Language and Communication Research at Cardiff University. Her specialism is in corpus-based critical discourse analysis of representations of identity in texts and topics in public and professional communication, most recently: (social) media discourse, medical communication, and language of the law.

Digital trust, drugs and the extreme far right: Too much for CADS?

Nuria Lorenzo-Dus
Swansea University
n.lorenzo-dus@swansea.ac.uk

Matteo Di Cristofaro
Swansea University
m.dicristofaro@swansea.ac.uk

In this presentation we will describe a research programme that uses CADS as an interdisciplinary methodology in order to explore one of the most widely debated and complex questions of our digital society: how do individuals and groups seek to generate trust in digital environments? Two case studies are selected to this end. The first is selling and buying drugs in the Dark Net, specifically in the flagship crypto-drug market Silk Road. The second is promoting extreme far-right values via Facebook and Twitter by the groups Reclaim Australia and Britain First. Whilst clearly different in terms of the activities performed and the digital platforms used, both case studies centrally involve attempts at generating trust discursively. In crypto-drug markets, users seek to enhance their reputation within a highly competitive, extra-judicial environment by, for instance, offering advice about avoiding being scammed by other users. Similarly, the extreme far right groups Britain First and Reclaim Australia use social media posts to mobilise a range of themes across domains as varied as food certification and vaccines in order to inculcate and create a sense of trustworthiness in their values, while delegitimising those of their perceived enemies.

In addition to presenting the key results of our case studies, our talk will include a (self-) critical reflection of what working with digital corpora entails from a technical and analytic perspective. Language data from digital environments are tied to the 'structure' of the platform through which they are published on the web. Creating a corpus of such data involves understanding this structure, which differs from platform to platform and which has a direct influence on both how the corpus tools can be used and on the language analysis itself. Through our case studies, we will reflect on how corpus linguistics methodologies may overcome the complexities of digital datasets as well as the current limitations faced. Finally, our presentation will reflect upon the challenges and opportunities of integrating CADS methods and policy-making and law enforcement practical needs.

Nuria Lorenzo-Dus is Dean of Postgraduate Research and Professor of Applied Linguistics at Swansea University. Her research uses interdisciplinary research methods that keep CADS at its core. Her current research examines digital discourses of trust and influence (ab)use, focusing on illegal contexts such as child sexual exploitation, crypto-drug markets and cyber-terrorism.

Matteo Di Cristofaro is a researcher at Swansea University (College of Law and Criminology). His main areas of expertise are Corpus Linguistics and Cognitive Linguistics, which he employs to develop both academic and non-academic application of linguistic approaches, in particular Corpus-Assisted Discourse Studies. He has applied these methodologies to the analysis of Far-Right propaganda on social media, and of radicalisation narratives in Islamic extremist groups online magazines. In the linguistics field he is mostly interested in cognition and language change: He has recently worked on the analysis of cognitive processes (Resonance, Entrenchment Inhibition) involved in language creativity, and is currently researching on the diachronic evolution of epistemicity and evidentiality in Italian.

Blind Spots and Dusty Corners: (self)-reflections on partiality in corpus & discourse studies

Charlotte Taylor
University of Sussex
charlotte.taylor@sussex.ac.uk

Anna Marchi
University of Bologna)
anna.marchi@unibo.it

In this talk we address the issue of partiality in corpus & discourse studies. The combination of corpus linguistics and discourse analysis, in its varied forms, has become an established research enterprise, as celebrated in this conference. One of the assumed added values of this match is that greater objectivity is achieved through the use of corpora and corpus tools. However we find the focus on objectivity problematic: each and every step of the analysis affects the outcome of research and it seems relevant, as C&D studies come of age, to revisit our methodological and epistemological assumptions. The partiality we discuss refers to both the potential incompleteness of research and to potential bias or influence: the 'dusty corners' and the 'blind spots' of the title. Dusty corners are the neglected aspects of analysis, for instance the tendency to ignore what is absent in the data or the emphasis on differences at the expense of similarities. The blind spots are those aspects that, by being unseen, may inadvertently impact on the results, for example data segmentation, visualisation or choice of metrics. We're not setting out to provide all the answers, but to raise a series of questions and provoke greater awareness of the implications of methodological choices and research practices. Since all research, no matter whether it is framed as theory-driven or data-driven, is necessarily and foremost researcher-driven, we argue that self-reflexivity should be an essential aspect of our work.

Charlotte Taylor is Senior Lecturer in English Language & Linguistics at the University of Sussex. Her research centres around language and conflict and she is particularly interested in theoretical-methodological issues of how we combine (critical) discourse studies and corpus linguistics. Her publications include Corpus approaches to discourse: a critical review (2018, co-edited with Anna Marchi), Exploring Silence and Absence in Discourse: Empirical Approaches (2018, co-edited with Melani Schroeter), Mock Politeness in English and Italian: A corpus-assisted metalanguage analysis (2016), and Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (2013, co-authored with Alan Partington & Alison Duguid).

Anna Marchi is an adjunct lecturer at the University of Bologna and collaborated with the Universities of Siena, Cardiff, Swansea and Lancaster. Her research interests are in the area of methodology of corpus linguistics and discourse analysis, news discourse and linguistic approaches to journalism studies. She is currently preparing a monograph for Routledge: Self-reflexive Journalism: A Corpus Study of Journalistic Culture and Community in The Guardian.

CONFERENCE PAPERS

Ordered alphabetically by first author's LAST name

Attitude markers in political speeches: a cross-cultural corpus-based

Ana Albalat-Mascarell
Universitat Politècnica de València
analmas@idm.upv.es

Within Hyland's (2005) metadiscoursal framework, attitude markers are common rhetorical strategies that indicate the speaker's affective, rather than epistemic, attitude to propositions. By the frequent use of evaluative terms signalling surprise, agreement, importance, and so on, speakers manage to build a close intimacy with their interlocutors with whom they share their own feelings and emotions (Crismore and Farnsworth, 1989; Hyland, 2005, 2010, 2017; Hyland and Tse, 2004; Mur-Dueñas, 2010). These signals of affect play a key role in strengthening the persuasiveness of political messages by decreasing the level of formality and showing greater understanding of the perspectives and problems of the audiences (Reyes, 2015). Attitude markers are therefore one of the main means by which politicians create rapport with the public, bringing familiarity to speech events and projecting the image of an electorally viable persona (Lempert and Silverstein, 2012).

In this vein, this paper aims at quantitatively and qualitatively analysing the presence and functions of attitude markers in a comparable corpus of North-American and Spanish political speeches. My objectives are, first, to examine the frequencies of the words and phrases performing an affective role in campaign speeches aimed at a large audience; second, to compare the rhetorical function of the most frequently used expressions and relate them to the candidates' persuasive aims; third, to explore cross-cultural differences regarding the distribution and rhetorical roles of these strategies and contrast them with the outcome of each election. In the methodology set for this study, the analysis was based on a corpus of authentic data consisting of the transcripts of campaign speeches involving the leaders of at least the two parties topping opinion polls in each country and election (i.e. the PP and the PSOE (also Podemos in the 2016 election) in Spain and the Democratic and Republican parties in the United States). Furthermore, the quantitative use of attitude markers was analysed with the tool 'Metool' developed specifically to detect metadiscourse strategies.

Overall, results indicate a similar frequency of use and tendencies in the rhetorical purposes of attitude markers in both sub-corpora. Moreover, the people, entities and events which are most often attitudinally evaluated are also similar: opponents and political enemies, own candidacies and campaigns, elections, policies the speaker implemented in the past, policies the speaker promises to support when elected, and policies adopted by previous governments and political

adversaries. It could be concluded that politicians share a cross-culturally common set of rhetorical devices used to emphasize the speaker's attitude towards certain preferred topics and establish a connection with the audiences. However, differences in the use of attitude markers between the political actors involved in each national group suggest a possible correlation with the speaker's communicative characteristics and his or her perceived political persona.

References

Crismore, A., Farnsworth, R., 1989. Mr. Darwin and his readers: exploring interpersonal metadiscourse as a dimension of ethos. *Rhetoric Review*, 8(1), 91-112.

Hyland, K., 2005. *Metadiscourse*. London: Continuum.

Hyland, K., 2010. Metadiscourse: Mapping interactions in Academic Writing. *Nordic Journal of English Studies*, 9(2), 125-143.

Hyland, K., 2017. Metadiscourse. What is it and where is it going? *Journal of Pragmatics*, 113, 16-29.

Hyland, K., Tse, P., 2004. Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25(2), 156-177.

Lempert, M., Silverstein, M., 2012. *Creatures of Politics: Media, Message, and the American Presidency*. Indiana: Indiana University Press.

Mur-Dueñas, P., 2010. Attitude markers in business management research articles: a cross-cultural corpus-driven approach. *International Journal of Applied Linguistics*, 20(1), 50-72.

Reyes, A., 2015. Building intimacy through linguistic choices, text structure and voices in political discourse. *Language & Communication*, 43, 58-71.

A Corpus Linguistic Analysis of UK Reuters Online Coverage on Syrian Revolution

Amaal Algamde
Bangor University
Elp45f@bangor.ac.uk

The Syrian revolution is one of a series of revolutions that erupted in the Middle East in 2011 under the name of Arab Spring. This incident became a major issue, which attracted the attention of international mass media and led to the involvement of global super powers. The research is premised on the assumption that media discourse is biased and imbued by political, economic and social contexts (Fowler, 1991). Based on this assumption, the study addresses the discursive constructions of anti- and pro-government powers in the Syrian revolution that appear in a 1,000,000-word corpus of UK Reuters' online reports, issued between 2013 and 2015. The accessibility of digital resources and heavy global reliance on online news validate the significant role online news plays in shaping and influencing the attitudes of potential readers. The study focus on the Syrian revolution in particular contributes to the bulk of existing literature on Arab Spring studies and motivates perceptions on the ideologies underpinning the discourses of news media. This study's examination of UK Reuters news coverage is the first of a twofold comparative study tackling representations of the Syrian revolution in UK and Middle East online news.

This study adopts a corpus assisted discourse analysis approach to investigate Reuter's ideology in reporting the events that occurred during the Syrian conflict. The analysis examines three types of lexicosemantic relations: the semantic macrostructures within which the two social actors are framed, the lexical collocations characterising the news discourse, and the discourse prosodies they convey about the two sides of the conflict. The study utilises two computer-based approaches, the Sketch engine (Kilgarriff et al, 2004) and Antony's (2014) AntConc

software, to minimize the bias of subjective analysis. Conducting a bottom-up analysis, the study first extracts the keywords list to reflect the news agency's focus in relation to the reference corpus, which reveals a pro-Syrian regime stance. The keywords are then further analysed with the lexical cluster tool and concordances to explore the common lexical patterns and distinct semantic macrostructures within which the two conflicting sides are introduced. The collocational patterns characterising the social actors (SAs) are identified and ordered according to their salience, based on log-dice scores. The analysis further accounts for three colligation patterns, namely the subject of verbs, the object of verbs and modifiers, using Sketch engine. In the light of the given representations, the study points out the underlying discourse prosodies, which reflect positive and negative representations of SAs. Finally, the ideological implications are discussed within the socio-political context of the Syrian revolution.

The findings show a kind of stereotypical representation of anti- and pro-government SAs. Reuters introduces the Syrian revolt within a sectarian frame. The SAs are commonly presented according to their religious background as Sunnis and Shi'ites rather than their political orientation. By and large, the government SAs are negatively introduced. They are thematically framed within the contexts of gassing civilians, cracking down on demonstrators, refusing an authorised transition and international condemnations government's violence, while the anti-government SAs are given dual representations, which are ideologically motivated. Reuters presents two labels for anti-government SAs: Islamists and rebels. Islamists are represented as radicals hiding a political agenda and are always embedded in a context of threat and violence. On the other hand, rebels, including the opposition, are positively represented. They are characterised in terms of their resistance and power and introduced as moderates who are eligible for military support. The examination of lexical patterns and concordances reveals incidents, which gained a great focus in Reuters such as, the government chemical attacks against civilians, the threat of Islamists dominance after Assad's fall and the issue of the government's suppression of Sunnis since 1971. These results can be used as a reference analysis to facilitate a more in-depth CDA analysis of media coverage on the Syrian revolution. In the context of this study, a comparison between CL findings on the Syrian revolution and other Arab revolutions' CL results would offer insight on how the Arab Spring has been globally reported.

References

Anthony, L. 2014. AntConc (Version 3.4. 4) Wasada University: Tokyo, Japan.
Fowler, R. 2013. *Language in the News: Discourse and Ideology in the Press*. Routledge.
Kilgarriff, A., Rychly, P., Smrz, P., and Tugwell, D. 2004. "Itri-04-08 the Sketch engine". *Information Technology*, *105*. 116.

Semantic Prosody in Nature as a Qur'anic Theme

Hala Alshahrani
Princess Nourah bint Abdulrahman
University /University of Leeds
mlhasl@leeds.ac.uk

Claire Brierley
University of Leeds
c.brierley@leeds.ac.uk

Semantic prosody is defined as the spreading of connotational colouring (Partington 1998, 2004); it refers to lexical items that collocate with semantic classes of words that are positive;

negative; or neutral. This tendency of words to occur in a certain semantic environment is one of the various linguistic features which gives the Qur'anic text harmony and cohesion. There has been little work done on collocation and semantic prosody in the Quran. Accordingly, this study participates in shedding light on the collocational behavior of natural terms, such as earth; day; sky; heaven; garden; mountain; etc., as they appear in different environments. We used the Python based Natural Language Toolkit (Bird et al 2009) and corpus linguistics techniques to examine the semantic behavior and the implied attitudinal meanings of a nature in the Quran. We explored the semantic prosody of nature as a Qur'anic theme (Rahman 2009) via a list of natural phenomena terms over Yusuf Ali's acclaimed English translation available from the Tanzil project[1]. We ascertained statistically significant collocational patterning based on the n-best list rule of the highest ranking candidates (Evert 2005). Our first task was to generate frequency distributions of the natural phenomena terms over the corpus. Based on their frequencies and the contexts in which the terms occur, we developed a taxonomy for textual content analysis to produce conceptual categorization of the Qur'anic contextual environments of nature, such as: glorifying God; creation of God; punishment; Day of Judgment calamities; etc. Our next step was to extract raw frequencies for bigrams, and then verify true collocations via application of Log Likelihood metric of collocation extraction. The choice of the statistically significant collocational patterning was based on the *n-best list* rule of the highest ranking candidates (Evert 2005). After we examined the 20-best bigrams for each of the terms in the light of their contextual environments, we labelled each with a connotational colouring or Semantic Prosody (SP) (as in Sinclair 2003: *positive*; *negative*; and *neutral*). One interesting finding was that nature in its positive Quranic connotation semantically denotes landscape and cosmos (e.g. *earth; moon; rivers; gardens; rivers*; *etc.*), while in its negative SP, it denotes weather (e.g. *thunder; lightning; shower; flood: etc.*). Finally, we analysed the positive, negative or neutral valence of verified collocations for insight into the symbolism of nature in the Quran. Our findings show that the majority of semantic prosody of nature in the Quran is mostly positive. The highest number of verses in which nature appears in the Quran relies on concepts of God's creation; glorifying God; reward; and description of believers. In future studies, each instance of SP pertaining to natural phenomena can be further explored in depth to relate them to pragmatic functions of nature as a Quranic theme.

References

Bird, S., Klein, E., Loper, E (2009). Natural Language Processing with Python. O'Reilly Media Inc. ISBN 0-596-51649-5 .

Evert, S. & Kermes, H. (2003). Experiments on candidate data for collocation extraction. In Companion Volume to the Proceedings of the 10th Conference of The European Chapter of the Association for Computational Linguistics, pages 83–86

Evert, S. (2004). The Statistics of Word Cooccurrences: Word Pairs and Collocations. Dissertation, InstitutfürmaschinelleSprachverarbeitung, University of Stuttgart. Published in 2005, URN urn:nbn:de:bsz:93-opus-23714. Available from: http://elib.uni-stuttgart.de/opus/volltexte/2005/2371 ./

Partington, A. (1998). Patterns and Meanings - using corpora for English language research and teaching. Amsterdam: Philadelphia: John Benjamins.

Partington, A. (2004). "Utterly content in each other's company": Semantic prosody and semantic preferences. International Journal of Corpus Linguistics, 9(1), 131-156.

Python 2.7.Tutorial. (2015) https://www.youtube.com/user/sentdex [October 2016]

Rahman, F. (2009).Major Themes of the Quran (2nd ed.), with a new Forwarded by EbrahimMoosa. Chicago: IL. University of Chicago.

Sinclair, J (2003). Reading Concordances. London: Longman

---

[1] http://tanzil.net/ :is a web resource consisting of the original verses in Arabic as well as 42 manual translations of the whole book.

An Intersectional Approach to Saudi Women Inclusive Identity in BBC News Discourse: A Corpus-assisted Critical Study

Zainah Alshahrani
Cardiff University
AlshahraniZT@cardiff.ac.uk

How is the inclusive identity of Saudi women (SW) constructed in BBC news discourse? How does this reflect on the way they are perceived? What stereotypical gendered ideologies and hegemonic conceptions are reproduced in the examined data? Within the framework of *intersectionality* these questions are addressed to see how intersects of social dimensions, as constructing phenomena, shape complex identities. Intersects of social categories in the Saudi as well as in the Other's institutions are combined and work reciprocally to shape what Collins (2000) call 'dominance matrix' which constructs a negative disadvantaged identity for Eastern females in the Western news context. I adopt feminist critical discourse analysis (FCDA) in this study to highlight what leads to the discursive practices of trivialization in news language, as a social practice, as expressed by some language features and strategies, the thing that has its impact on how the identity of 'women' who are 'Saudi' is constructed and consequently negatively perceived. The 'theoretical eclecticism' of FCDA makes it elastic to adopt intersectionality theories that underlie both individuals' multiple identities and the social practices against them (Lazar, 2007: 183). Thus, the framework recognises that systems of power exist within society to reinforce oppressions and or privileges that influence identity (systemic intersectionality); and secondly that women are comprised of different identity features that are shaped by the power system of the social institution 'constructionalist intersectionality' (Prins 2006). The data encompasses 52 texts that mount to 35000 words generated from BBC news website. Each text should include the search terms 'Saudi wom*n/Saudi female(s)' and should be broadcasted during the period between 2008 and 2017. The data was then uploaded to Sketch Engine corpus tool. The analysis is conducted quantitatively and qualitatively integrating some corpus methods with some discourse analytical tools. The corpus analysis of both *keyness* and *collocates* that associate women, in addition to sketching the female gendered terms help to explore the different sides of the multiple identity which are shaped by intersects of social dimensions. It also helps to identify the semantic themes which are associated with women. The qualitative analysis of *concordance* lines and some chosen texts broadcasted in different points of time help to explore how the lexico-structural language choices were manipulated to position women as passive social actors in the processes and as recipients of hegemonic practices. The findings reveal that the intersectional dimensions of policy (rules imposed on women), culture (male dominance), religion (Islamic doctrine) and gender (women subordination) and from the other's side, race, religion and gender (being an eastern Muslim women) are combined in BBC news discourse to construct a weak, submissive, dominated and passive identity for Saudi women. In addition, women are shown as being involved in semantic themes that are highly connected to such systems of power. Though the examined period witnessed sequence of positive changes in women's status and their social roles, a reproduction for the stereotypical gendered images for Saudi women is shown by the linguistic construction of their identity. However, there is a clear shift in the semantic themes which at the end of the examined period (2017) focused overtly on women's appearance and the way they are dressed. The results confirm the intrinsic role of analysing different intersectional social aspects to understand and unpack the complex identity in a

specific social institution. In other words, it shows the importance of using an intersectional approach to conduct a critical study. This study is an attempt to add a theoretical and methodological contribution to women's critical studies in news discourse and it is likely to raise the awareness towards other purposes of practices of gender bias and practices against women in different social institutions.

References

Collins, P. 2000. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment.* New York: Routledge.

Crenshaw, K. 1991. Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review* 43 (6), pp 1241- 99.

Lazar, M. 2007. Feminist critical discourse analysis: Articulating a feminist discourse praxis. *Critical Discourse Studies* 4(2), pp.141-64.

Prins, B. 2006. Narrative accounts of origins: a blind spot in the intersectional approach. *European Journal of Women's Studies* 13 (3), pp 277-90.

Metaphors in U.S. Same-Sex Marriage Discourse: Toward an Integration of Corpus Linguistics, Cognitive Linguistics, and Critical Discourse Analysis

Pongbodin Amarinthnukrowh
Department of Linguistics, Chulalongkorn University
pongbodin.a@outlook.com

Through an integration of the corpus, the cognitive, and the critical discourse approaches, this study aims to investigate metaphorical expressions used in web-based news articles regarding the legalization of same-sex marriage in the United States. As same-sex marriage was ruled to be legal nationwide on 26 June 2015, news articles from January to June 2015 were extracted from the NOW (News on the Web) corpus (Davies 2013). Upon the compilation of a corpus, the articles extracted were uploaded to Antconc (Anthony 2014) and sorted through for relevant data, resulting in a corpus containing 172 articles of 146,029 words. This corpus was then uploaded to Wmatrix (Rayson 2008) to annotate the corpus at part-of-speech and semantic levels. Subsequently, the purpose-built corpus was compared to the BNC sampler written informative corpus (745,726 words) with the log-likelihood (LL) cut-off point at 15.13 for 99.99% significance to find statistically significant semantic domains. Top five domains achieved by this means are: 'relationship: intimacy and sex' (LL 6230.34), 'kin' (LL 5090.29), 'law and order' (LL 3643.84), 'government' (LL 1453.62), and 'religion and the supernatural' (LL 1249.32). However, at this point, I can present only the analysis of the first statistically significant tag as this study is part of my ongoing M.A.'s Thesis.

Based on the Metaphor Identification Procedure (Pragglejaz Group 2007), metaphorical expressions in the 'relationship: intimacy and sex' tag were identified. To cover both conventional and novel metaphorical expressions, a mixed, adaptive method of metaphor analysis combining 'Conceptual Metaphor Theory' (CMT) with 'Conceptual Blending Theory' (BT) was employed. Conventional metaphorical expressions, on the one hand, were examined following the CMT traditions. Novel metaphorical expressions, on the other hand, were investigated by using BT.

Central to U.S. same-sex marriage discourse is the WAR/CONFLICT metaphor which was found

to have some conceptual links to other conceptual metaphors, namely the GAMES, the DISEASE/ILLNESS, and the JOURNEY metaphors. Also, the metaphorical expressions in this dataset, especially the novel ones, confirm Dancygier and Sweetser (2014)'s claim that metaphors are "viewpointed". This is because text-producers normally coin creative and novel terms to suit the purposes of their communicative intent. Thus, by analyzing both conventional and novel metaphorical expressions, clearer pictures of how the same-sex marriage controversy was constructed, reinforced and/or attenuated can be obtained.

References

Anthony, L. 2014. AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available at http://www.laurenceanthony.net/.

Dancygier, B. and Sweetser, E. 2014. *Figurative Language*. Cambridge: Cambridge University Press.

Davies, M. 2013. *Corpus of News on the Web (NOW): 3+ billion words from 20 countries, updated every day*. Available at https://corpus.byu.edu/now/.

Pragglejaz Group. 2007. "MIP: a method for identifying metaphorically used words in discourse". *Metaphor and Symbol* 22 (1), 1-39.

Rayson, P. 2008. "From key words to key semantic domains". *International Journal of Corpus Linguistics* 13 (4), 519-549.

Which techniques of down-sampling best complement a corpus-assisted discourse analysis? A case study on press representations of obesity

Paul Baker
Lancaster University
j.p.baker@lancaster.ac.uk

Baker and Levon (2015) have argued that corpus assisted discourse analysis can work well as part of a triangulation approach involving close reading of a down-sampled set of corpus texts. However, a question arises regarding which method(s) of down-sampling are most likely to reproduce or complement the findings of the corpus analysis.

In this paper, media reporting of obesity is used as a case study in order to compare traditional 'close reading' methods of discourse analysis with a method of analysis commonly used in corpus assisted discourse studies – concordance analyses of relevant collocational pairs in a corpus. Media reporting of obesity has been criticised in academic research as alarmist and uncritical (Holland et al 2011), and is perceived by obese people as portraying them as freaks and enemies of society who are rarely given a voice unless successfully losing weight, which Couch et al (2015) argue is a form of 'synoptical' social control. In reporting on obesity, Caulfield et al (2009) claim that newspapers can influence perceptions which has implications for public policy e.g. their diachronic study of US newspapers indicated a shift from a deterministic view of obesity (e.g. genetic factors) towards one based on personal responsibility (e.g. diet and exercise).

This study examines a corpus of British newspaper articles about obesity (2010-2015), collected from LexisNexis, with the aim of a) identifying how language is used to represent obese people and b) comparing traditional close reading with a corpus-based approach to discourse analysis. Four down-sampling techniques were used in order to identify sets of 10 articles for close reading. These were 1) sampling articles from the week where the highest number of articles

were published, 2) sampling articles that contain the highest number of references to obesity 3) random sampling and 4) sampling based on using the tool ProtAnt which ranks the proto-typicality of articles based on the number of keywords found in them. The close reading considered phenomena such as quotation patterns, narrative structure, argumentation strategies and fallacies as well as lexical choice, grammatical relationships and metaphor. The analyses considered the following research questions:

1. What types of people (or other social actors) are particularly represented as obese or affected by obesity?
2. How are such people represented, particularly in terms of evaluations?
3. What is represented as the consequences of obesity?
4. What is represented as the cause of obesity, and what actions are suggested in order to reduce obesity?
5. What strategies are used in order to legitimate the positions taken in response to the above questions?

Additionally, collocates of the terms *obese* and *obesity* were identified, grouped into semantic categories, and then concordance lines of a range of collocates taken from different categories were analysed in order to address the above research questions. To ensure a degree of comparability across the different analytical conditions, the same amounts of time were spent on each form of analysis.

Having carried out the analyses, and identifying findings a meta-analysis compared the findings elicited by different techniques in order to identify the extent that they overlap or give dissonant results. Rather than attempting to judge which approach was the most successful, the paper ends with a more reflective discussion of their strengths and weaknesses and makes suggestions for how they can be combined in order to complement one another.

References

Baker, P. and Levon, E. (2015) 'Picking the right cherries?: a comparison of corpus-based and qualitative analyses of news articles about masculinity.' *Discourse and Communication* 9(2): 221-336.

Caulfield, T. Alfonson, V. and Shelley, J.(2009) Deterministic?: Newspaper representations of obesity and genetics. *The Open Obesity Journal* 1: 38-40.

Couch, D., Thomas, S. L., Lewis, S., Blood, R. W.and Komesaroff, P. (2015) Obese adult's perception of news reporting on obesity. The panopticon and synopticon at work. Sage Open 5(4) 2158244015612522.

Holland K., Blood R. W., Thomas S. I., Lewis S., Komesaroff P. A. and Castle D. J. *(2011). "Our girth is plain to see": An analysis of newspaper coverage of Australia's future "Fat Bomb." Health, Risk & Society, 13, 31-46.*

Cross-linguistic corpus-assisted discourse studies: Identifying similarities and differences by comparing thematic categories of words across languages

Paul Baker
Lancaster University
j.p.baker@lancs.ac.uk

Rachelle Vessey
Birkbeck, University of London
r.vessey@bbk.ac.uk

This paper uses corpus linguistics and qualitative, manual analysis to compare corpora of English and French Islamist extremist texts. Drawing on 679,743 words in English and 191,344

words in French, we use AntConc (Version 3.4.4) (Anthony 2017) to examine the extent to which extremist messages in each language draw upon similar and distinct themes and linguistic strategies. Given that the data exist in different languages, a direct comparison of the datasets is not possible. Therefore, this paper discusses our innovative methodology: a cross-linguistic corpus-assisted comparison of thematic categories.

We compiled English and French keyword lists using the BE06 (Baker 2009) as an English reference corpus and 1,462,398 words of La Presse newspaper articles as a French reference corpus. The 500 top-ranked English and French keywords were examined using collocate, cluster and concordance analysis. Next, each keyword list was examined for themes that emerged from the use of words in context. Themes that emerged in one language were compared against the themes that emerged in the other language. Then, the English and French keyword lists were compared in order to establish which keywords were common to both corpora and which were unique to the English and French data. A final step involved mapping the equivalence (or non-equivalence) of keywords across languages (i.e. the cross-linguistic similarities and differences) onto the thematic categorisations previously established. This required revisiting the themes established in the previous step and also, in some cases, revisiting the qualitative analysis and revising some of the categories. The ultimate outcome of this procedure was a matrix that cross-listed the emergent thematic categories against the keywords that were either common to both languages or unique to one set of language data. As a final step, we assessed the similarities and differences across languages by calculating the sum of relative frequencies of all words in each category and the differences between these relative frequencies in English and French. Where relative frequencies were both high and similar in English and French, we consider these to be shared themes. Where relative frequencies were exceptionally higher in one language category than the other (using %diff calculation, Gabrielatos and Marchi 2012), we consider these themes more salient to that language.

Our findings revealed numerous similarities as well as differences. Both corpora focus on religion and rewards (i.e. for faith) and strongly rely on othering strategies. However, the English texts appear to be more concerned with world events pertaining to Islam and the French texts focus on issues specific to France. Also, while the English texts draw on code-switching to Arabic as a form of legitimation, the French texts use a formal register and draw heavily on quotation from scripture in order to discuss permissions, rights, obligations and laws. Finally, the English texts refer to and justify violence to a greater extent than the French texts.

We argue that a comparison of thematic categories across languages is a useful way to identify how similar meanings can be expressed differently in different languages. However, we argue that the subtle nuances inherent to meaning-making require the manual creation of categories rather than automatic categorisation or the use of pre-existing schemes. Although the lack of similar and comparable corpus resources in both languages raise some limitations, we call for more work to be undertaken on French language corpus creation and analysis and also for more experimental studies across languages. Finally, we contend that our approach provides a new and innovative way to undertake cross-linguistic corpus-assisted discourse studies.

References

Anthony, L. 2017. *AntConc* (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baker, P. 2009. "The BE06 Corpus of British English and recent language change". *International Journal*

of Corpus Linguistics, 14:3, 312-337.

Gabrielatos, C. and Marchi, A. 2012. "Keyness: Appropriate metrics and practical issues". *Paper given at CADS 2012 University of Bologna,* 14 September 2012.

Using the word sketch tool to examine collocates of the lemma SCHIZOPHRENIC (n.) relating to 'dangerousness' in the UK press

James Balfour
Lancaster University
j.balfour@lancaster.ac.uk

A wealth of previous research has shown that people with schizophrenia are overrepresented as violent and dangerous in the UK press (e.g. Schizophrenia Commission Report, 2012; Clement Foster; Chopra and Doody, 2007) and that this contributes to stigmatising and misinformed public attitudes (Angermeyer et al, 2005). This is despite statistical evidence showing that people diagnosed with schizophrenia are not significantly more likely to be dangerous to others than members of the general public (Fazel and Grann, 2006, Walsh et al, 2002). Indeed, a US study showed that people with schizophrenia are 14 times more likely to be the targets of violence than the perpetrators of it (Brekke et al, 2001).

While several authors allude to the role of language and the manner in which the press draw associations between people with schizophrenia and violent crime (e.g. Kalucy et al, 2011; Thornicroft et al, 2015) no studies to date have seriously explored how these representations are linguistically encoded and sustained. In this conference paper, I use the word sketch tool, available via the online corpus toolbox Sketch Engine (Kilgarriff et al, 2014), to examine collocates of the lemma SCHIZOPHRENIC (n.) in a specialised 15 million word corpus of U.K. national press articles reporting on schizophrenia between 2000 and 2015. In doing so, I uncover groups of collocates in varying grammatical relationships with SCHIZOPHRENIC which cumulatively furnish the label with a semantic prosody relating to 'dangerousness'. Specifically, SCHIZOPHRENIC was found to co-occur unusually frequently with modifiers relating to perceived dangerousness and when in subject position, collocated with verbs referring to (sometimes quite lurid) violent actions. Furthermore, the label SCHIZOPHRENIC sometimes collocated with identities defined by their perceived dangerousness to others. More subtle patterns include collocates indexing lay-diagnoses which ascribe the label solely on the basis of violent sentiments or behaviour. I conclude by speculating on why the press may choose to represent schizophrenic people in such a biased and inaccurate way despite firm evidence to the contrary.

References

Angermeyer, M. C., Dietrich, S., Pott, D., & Matschinger, H. (2005). Media consumption and desire for social distance towards people with schizophrenia. *European Psychiatry, 20*(3), 246-250. doi:10.1016/j.eurpsy.2004.12.005

Brekke, J. S. (2001). Risks for Individuals With Schizophrenia Who Are Living in the Community. (Psychiatric Services).(Abstract). *JAMA, The Journal of the American Medical Association, 286*(23), 2922.

Chopra, A. K., & Doody, G. A. (2007). Schizophrenia, an Illness and a metaphor: Analysis of the use of the term 'schizophrenia' in the UK national newspapers. *Journal of the Royal Society of Medicine, 100*(9), 423-426. doi:10.1177/014107680710000919

Clement, S., & Foster, N. (2008). Newspaper reporting on schizophrenia: A content analysis of five

national newspapers at two time points. *Schizophrenia Research, 98*(1), 178-183. doi:10.1016/j.schres.2007.09.028

Fazel, S., & Grann, M. (2006). The population impact of severe mental illness on violent crime.(Author abstract). *American Journal of Psychiatry, 163*(8), 1397. doi:10.1176/ajp.2006.163.8.1397

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovvář, V., Michelfeit, J., Rychlý, P., Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, *1*(7-36). http://www.sketchengine.co.uk

Thornicroft, G., Rose, D., Evans-Lacko, S., Corker, E., Goulden, R., & Henderson, C. (2011). Newspaper coverage of mental illness in the UK, 1992-2008. *BMC Public Health, 11*(1), 796. doi:10.1186/1471-2458-11-796

Walsh, E., Buchanan, A., & Fahy, T. (2002). Violence and schizophrenia: examining the evidence. *The British jour*

A corpus-based study on the progressive use of stative verbs

Serap Atasever Belli
Anadolu University
serapatasever@anadolu.edu.tr

Ümit Deniz Turan
Anadolu University
udturan@anadolu.edu.tr

An insight into the corpus of present-day language provides a wealth of information about the typical language patterns and more accurate analysis of language in use that best illustrates natural discourse (Biber & Xeppen, 1998: 145). Stative verbs were traditionally assumed to be incompatible with the progressive aspect and even ungrammatical (i.e. Comrie, 1976; Kroeger, 2005). However, they have substantially increased in frequency and extended in functions in the progressive use in native British and American English (Mair, 2006; Aarts et al., 2010; Tumert, 2010; Dráčková, 2011; Granath & Wherrity, 2014, among others). However, it has not been extensively explored whether there has been a difference or recent change in the frequency of progressive use of stative verbs from 1990 until now and which functions stative verbs served in the progressive form across spoken and written varieties of language. So, the purpose of the present study is to investigate four semantic categories of stative verbs (i.e. perceiving, internal sensation, relationship/ a state of being and a state of mind/feeling) used in the progressive form in written and spoken English to explore both their frequencies and functions over a period of time from 1990 to the present. The following research questions were addressed in this study:

1. What is the frequency of distribution of stative verbs used with the progressive?
   a. Across five genres (spoken, academic, newspaper, magazine, fiction)
   b. Across different time periods (1990-2017)
   c. Across four sub-categories of stative verbs
2. What are the functions of stative verbs when used with the progressive?

For the purposes of the study, 12 stative verbs (*i.e. love, know, remember, see, hear, feel, ache, hurt, taste, belong, resemble, cost)* chosen from four categories based on Leech and Startvik's (2002:124) classification, were analysed in *The Corpus of Contemporary American English* (COCA). COCA is the largest freely-available corpus of English including over 560 million words dating back to 1990 up to 2017, obtained from a variety of sources (i.e. academic texts, fiction, newspapers, popular magazines and spoken). During data analysis procedure, stative verbs used in the progressive form were extracted from the corpus by using a set of tags listed by Das (2010) in his study and all the instances were analysed both qualitatively for identification of progressive use and quantitatively for calculation of their frequencies. They were compared in terms of frequency of use across four categories and five registers and in different periods of time.

Overall, findings of the study yielded variation in the distribution of stative verbs used with the progressive across four categories and five genres. Stative verbs with the highest frequency of progressive use were found to be associated with perception and feeling verbs whereas those with the lowest frequency were related to state of being and cognition. Spoken and fiction genres included the highest frequency of progressive use; whereas, academic texts had the least frequency. Moreover, the results have led to substantial increase in the progressive use of certain stative verbs in recent decades.

References

Aarts, B., Close, J., and Wallis, S. 2010. Recent changes in the use of progressive construction in English. In B. Cappelle and N. Wada (eds.) *Distinctions in English grammar, offered to Renaat Declerk*. 148-168. Tokyo: Kaitakusha.

Biber, D., and Xeppen, R. 1998. Comparing native and learner perspectives on English grammar: A studv of complement clauses. *Learner English on computer.* 145-158.

COCA. Retrieved from https://corpus.byu.edu/coca/

Comrie, B. 1976. *Aspect, An introduction to the study of verbal aspect and related problems.* Cambridge: Cambridge University Press.

Das, D. 2010. "The Uses and Distribution of Non-Progressive Verbs in Progressive Forms in English: A Corpus-based Study". *Presented at the 26th Northwest Linguistics Conference.* Simon Fraser University, BC, Canada.

Dráčková, J. 2011. *Stative verbs with the progressive.* Unpublished MA thesis: Masaryk University, Czech Republic.

Granath, S., and Wherrity, M. 2014. "I'm loving you - and knowing it too: Aspect and so-called stative verbs". *Rhesis: Linguistics and philology, 4*(1), 2-22.

Kroeger, P. 2005. *Analyzing Grammar. An introduction.* Cambridge: Cambridge University Press.

Leech, G. and Svartvik, J. 2002. *A Communicative Grammar of English.* Third Edition. London: Longman.

Mair, C. 2006. *Twentieth-century English. History, Variation and Standardization.* Cambridge: Cambridge University Press.

Tumert, L. 2010. *Anti-progressive verbs in the progressive: Frequency changes across time and genres in American English.* Unpublished MA thesis: University of Oslo: Oslo, Norway.

Climate change and displacement: A diachronic corpus-assisted discourse analysis of the climate change-migration nexus in the press

Cinzia Bevitori
University of Bologna
cinzia.bevitori@unibo.it

Jane Helen Johnson
University of Bologna
janehelen.johnson@unibo.it

While climate change has gradually and crucially become a 'defining symbol of our collective relationship with the environment' (Boycoff 2011:1), an investigation of its role in complex, multi-causal phenomena of human mobility has emerged as a salient policy-making issue only in the latest years (e.g. Bettini et al. 2016, Bettini 2013, 2017, McAdam 2010). According to Nash (2017), the period between the Cancun negotiations of the United Nations Framework Convention on Climate Change in 2010 and the 2015 Paris talks is seen to represent a crucial moment in policy making as regards the climate change and migration nexus.

The purpose of the paper is to explore representations of migration and displacement in the context of anthropogenic climate change through a diachronic corpus-assisted discourse analytical perspective (Baker et al. 2008, Partington et al. 2013, Bevitori 2010, 2014), focussing on newspaper discourse. Despite the increasing role of social media, in fact, traditional media

still play a pivotal role in shaping public perception and opinion of science, policy and institutional actors (Baker et al. 2013). A diachronic, domain-specific corpus of newspaper articles from a selection of UK and US broadsheets was gathered through the Nexis online searchable database. The corpus comprises all the articles containing the search words '*climate change*' AND '*migra**' OR '*refugee**' over three distinct 'critical' periods, selected diachronically to coincide with the Conferences of the Parties to the UN Convention on Climate Change, and specifically:

- between Cancun and Paris, (5 years - November 2010-2015);
- between Paris and Marrakech (1 year - November 2015-2016);
- since Marrakech, up to the present day (1 year - 2016-2017).

Some causes of migration are more publicly acceptable than others (Bank et al 2017); moreover a shift in thinking has been observed at policy-making level from 'climate refugees' to 'migration as adaptation' (Bettini et al 2016, Methmann and Oels 2015, Bose 2016) over these periods. Bearing in mind that migration as adaptation – with its associated baggage of 'resilience' ("the ability of a system and its component parts to anticipate, absorb, accommodate, or recover from the effects of a hazardous event in a timely and efficient manner" (IPCC 2012, 5)) – is also considered to be problematic (e.g. Clark and Bettini 2017, Rothe 2017), our investigation concerns the extent to which the mainstream press reflects this problematicity and particularly the extent to which these shifts in perception are associated with labels such as risk, threat and danger (e.g. Rolfe 2017, Johnson 2014): both to and caused by these displaced persons.

References

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyzanowski, M., McEnery, T. and Wodak, R. 2008, "A useful synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press". *Discourse and Society*, 19(3), pp. 273-306.

Baker, P., Gabrielatos, C. and McEnery, T. 2013. *Discourse analysis and media attitudes: the representation of Islam in the British press*. Cambridge University Press, Cambridge.

Bank, A., Fröhlich, C. and Schneiker, A. 2017. "The Political Dynamics of Human Mobility: Migration out of, as and into Violence". *Global Policy* Volume 8 . Supplement 1, pp. 12-18

Bettini, G., 2013. "Climate Barbarians at the Gate? A critique of apocalyptic narratives on 'climate refugees'". *Geoforum* (45), pp. 63-72.

Bettini, G. 2017. "Where Next? Climate Change, Migration, and the (Bio) politics of Adaptation". *Global Policy*, 8 (1), pp. 33–39.

Bettini, G., Nash, S.L., and Gnoli, G. 2016. "One step forward, two steps back? The fading contours of (in)justice in competing discourses on climate migration". *The Geographical Journal*, 183 (4), pp. 348-358.

Bevitori, C. 2010. *Representations of Climate Change. News and opinion discourse in UK and US quality press: a Corpus-Assisted Discourse Study*. Bologna: Bononia University Press

Bevitori, C. 2014. "Values, Assumptions and Beliefs in British Newspaper Editorial Coverage of Climate Change. In C. Hart and P. Cap (eds.) *Contemporary Critical Discourse Studies*. London: Bloomsbury Academic, pp. .623-625.

Bose, P. S. 2016. "Vulnerabilities and displacements: adaptation and mitigation to climate change as a new development mantra". *Area*, 48 (2), pp. 168–175.

Boykoff, M. T. 2011. *Who Speaks for the Climate? Making Sense of Media Reporting of Climate Change*. Cambridge : Cambridge University Press.

Clark, N. and Bettini, G. 2017. "'Floods' of Migrants, Flows of Care: Between Climate Displacement and Global Care Chains". In N. Gill, V. Singleton and C. Waterton (eds.) *The Sociological Review Monographs Special Issue*: `Care and Policy Practices'. Vol 65 (2), pp. 36-54.

IPCC. 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. *Special report of the International Panel on Climate Change*. Cambridge: Cambridge University Press.

Johnson, J. H. 2014 "Family in the UK – Risks, Threats and Dangers: A Modern Diachronic Corpus-assisted Study across Two Genres". In M. Gotti and D.S. Giannoni (eds.) *Corpus Analysis for Descriptive and Pedagogic Purposes: ESP Perspectives.* Bern: Peter Lang, pp. 263-285.

McAdam, J. (ed.). 2010. *Climate Change and Displacement: Multidisciplinary Perspectives.* London and Oregon: Hart Publishing.

Methmann, C. and Oels, A. 2015. "From 'fearing' to 'empowering' climate refugees: governing climate-induced migration in the name of resilience". *Security Dialogue* 46, 51–68.

Nash, S.L., 2017. "From Cancun to Paris: An Era of Policy Making on Climate Change and Migration". *Global Policy* DOI: 10.1111/1758-5899.12502.

Partington, A., Duguid, A. and Taylor, C. 2013. *Patterns and meanings in discourse. Theory and practice in Corpus-Assisted Discourse Studies.* Amsterdam/Philadelphia: John Benjamins.

Rothe, D. 2017. "Gendering Resilience: Myths and Stereotypes in the Discourse on Climate-induced Migration". *Global Policy* 8 (S1), 40–47.

Statistical choices in discourse analysis: How to efficiently analyse and visualise discourse data?

Vaclav Brezina
Lancaster University
v.brezina@lancaster.ac.uk

Statistical analysis of discourse involves an inherent paradox. While discourse is often fluid, ambiguous and fuzzy, statistics expects rigour, precision and clearly defined categories. So how do we reconcile this tension in statistical analysis of discourse? This methodological presentation engages with this fundamental question, offering a critical reflection on the state of the art in corpus-based discourse studies and providing suggestions for more efficient and rigorous analysis of discourse data (cf. Brezina 2018a and b). The presentation discusses (often implicit) assumptions connected with the application of mathematical models to the fluid reality of discourse and proposes a three-step procedure for dealing with discourse data. This procedure involves 1) the construct definition, 2) statistical operationalisation and 3) empirical analysis. Each step is clearly illustrated with multiple examples from a range of discourse studies.

The presentation also highlights some of the risks and temptations that statistical analysis of discourse involves. Quantitative analysis of language sometimes leads to reductionism, reducing semiotically rich data to surface forms easily analysable by automatic algorithms. This is especially apparent in some of the 'big data' approaches to language, which have recently emerged with the availability of large datasets ('big data'), which due to their sheer size cannot be explored manually (e.g. Michel et al. 2011). The presentation will show that the main points of departures of linguistically and socially informed analysis of discourse (e.g. Baker 2006, Partington et al. 2013) principally differ from the reductionist big data approaches, which pay little attention to linguistic and social contexts. Statistical analysis of discourse thus involves multiple choices at various levels. At the most general level of the 'philosophy' of analysis we need to choose an approach that best corresponds with the theoretical (linguistic) understanding of text and discourse (meaningful semiotic activity). The role of statistical analysis of discourse is to provide a lens through which the analyst can interpret, evaluate and contextualise the patterns observed in the data. This lens needs to be flexible enough to enable us to zoom out and see the larger picture as well as to zoom in on specific cases. This means that the statistical procedure does not replace the analyst but is used to focus their attention to important phenomena in discourse such as typical examples of language use, repeated associations and

points of semantic change. Returning to the paradox described at the beginning, we have to realise that the inherent fluidity, ambiguity and fuzziness of discourse offers a wonderful opportunity to explore the data from different perspectives using different methods; these perspectives can then be brought together via triangulation (Baker & Eckbert 2016), which provides a fuller picture of the complex linguistic and discourse reality than any single technique.

References

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Baker, P., and Egbert, J. (eds.) *Triangulating Methodological Approaches in Corpus Linguistic Research*. New York: Routledge.

Brezina, V. (2018a) *Statistics for corpus Linguistics: A practical guide*. Cambridge: Cambridge University Press.

Brezina, V. (2018b) In Taylor, C. & Marchi, A. (eds.) Statistical choices. *Corpus Approaches to Discourse: A Critical Review*. London: Routledge.

Partington, A., Duguid, A., & Taylor, C. (2013). *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies (CADS)* (Vol. 55). Amsterdam: John Benjamins.

A Killer, a Thief, a Tsunami: A Corpus-Assisted Study of UK Press Representations of Dementia

Gavin Brookes
University of Nottingham
gavin.brookes@nottingham.ac.uk

Dementia is a syndrome that causes a range of cognitive impairment symptoms, including problems with memory, reasoning, perception and communication. Dementia presents one of the biggest global public health challenges today, with 47.5 million people living with the syndrome worldwide, and 7.7 million new cases each year (WHO, 2016). In 2016, the British Office of National Statistics (ONS, 2016) published a report showing that dementia had now replaced cancer and heart disease as the leading cause of death in England and Wales, the syndrome now accounting for over 11 per cent of all deaths registered in 2015. The study presented in this talk examines contemporary representations of dementia in the British print media. The analysis takes a corpus-assisted approach to discourse analysis (Baker, 2006), based on a corpus of over 8,000 national press articles (approx. 4 million words) published between 2007 and 2017. The analysis will expose the tendency for press representations of dementia to be consistently negative, stigmatising and – in some cases – misleading, relying in particular on a rhetoric of violence and suffering, all the while perpetuating a decidedly biomedical model of dementia which leaves limited scope for people to 'live well' with the syndrome. It is argued that such broadly loaded and sensationalist representations not only construe dementia as a direful and malicious disease, but also, crucially, obscure the personal and social contexts in which the syndrome is understood and experienced, not least by people with dementia themselves. This intensely lurid type of representation not only fails to address the ageist misinformation and common misunderstandings that all too commonly characterise the general public's understandings of dementia, but is also likely to exacerbate the stress and depression frequently experienced by people with dementia and their families.

References

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

Burt, D. B. 1999. "Dementia and Depression". In: M. P. Janicki and A. J. Dalton (Eds.), *Dementia, Aging, and Intellectual Disabilities: A Handbook*. New York: Hamilton. 198-216.

Office for National Statistics. 2016. *Statistical Bulletin: Deaths Registered in England and Wales (Series DR): 2015*. Accessed January 2017. Available at: https://www.ons.gov.uk/peoplepopulationand community/birthsdeathsandmarriages/ deaths/bulletins/deathsregisteredinenglandandwalesseriesdr/2015.

World Health Organisation. 2016. *Dementia: Fact Sheet*. Accessed January 2017. Available at: http://www.who.int/mediacentre/ factsheets/fs362/en/.

Overcoming the Obsession with Politicians: An Intercultural and Diachronic Examination of EP's (original and translated) discourse

María Calzada Pérez
University Jaume I
calzada@trad.uji.es

María del Mar Sánchez- Ramos
University of Alcalá
mar.sanchezr@uah.es

The European Parliament (EP) is an unbeatable example of a multilingual setting whose discourse has a vital impact upon the real world. MEPs from (still) 28 different countries and a handful of political groups have the responsibility to contribute to the governance of the EU. Since the Lisbon Treaty this impact is stronger than ever. Consequently, exploring corpora of proceedings from the EP (vis-à-vis equivalent chambers), the main aim of this paper, is bound to yield important information regarding the world we inhabit. In order to pursue this aim, the present paper sides with Wodak (2011: 7) in advocating that research:

> should go beyond studies critical of the present and aim at comparative analysis both in diachronic and intercultural terms so as to overcome the 'obsession' with politicians (i.e. to make not only the language of politicians but also the 'act of talking politics' the subject of study).

Accordingly, the paper proposed here is conducted along both the intercultural and diachronic axes. It resorts to the exploration of the European Comparable and Parallel Corpus Archive of Parliamentary Speeches (ECPC). The ECPC is a set of bilingual parliamentary corpora, which consists of 2004-2014's proceedings of the European Parliament (EP), the Spanish Congreso de los Diputados (CD) and the British House of Commons (HC). Compiled at the Universitat Jaume I (Spain), the Archive draws on work done in projects such as the OPUS open source parallel corpus (OPUS, Tiedemann 2009), the Translational English Corpus (TEC, Laviosa-Braithwaite 1993, Baker 1999) and the English Norwegian Parallel Corpus (ENPC, Johansson 2007). However, it incorporates contextual (sociolinguistic and sociocultural) and metalinguistic (i.e. speakers' status, gender, constituency, party affiliation, birth-date, birth-place, post, and institutional body and sub-body of representation) data, through XML annotation, that makes it unique.

Drawing on Corpus-Assisted Discourse Studies (CADS) (Partington, Duguid, & Taylor, 2013), the ECPC's structure and annotated format makes it possible to perform various kinds of intercultural comparisons. Among them, (1) EP's English and Spanish may be confronted with

the languages of equivalent national parliaments; (2) EP's original English and Spanish may be studied side by side its corresponding translation; and (3) sub-sets of language production may be inspected (according to the sociolinguistic and metalinguistic parameters stated above). The ECPC is also pre-processed to performed diachronic comparisons of the kind most favoured within Modern Diachronic Corpus Assisted Discourse Studies (MD-CADS). For their comparisons, CADS and MD-CADS resorts to either traditional tools (e.g., frequency indicators, keywords, concordances, clusters and collocations), or slightly more innovative items (e.g., detailed consistency, key keywords, associates, and lockwords).

Consequently, after establishing the (MD-)CADS-informed grounds upon which the paper stands, and after introducing the ECPC Archive and its represented context, a sample of possible corpus-assisted comparisons is presented together with their micro/macro significance. As a result, the analysis helps produce an incipient map of similarities and differences of the language use within the various corpora confronted and which results in the enumeration of some of the common features of the EP discourse and some of the most idiosyncratic traits of (original and translated) EP linguistic behaviour.

References

Johansson, S. 2007. *Seeing through Multilingual Corpora on the Use of Corpora in Contrastive Studies*. Amsterdam/Philadelphia: Benjamins.

Laviosa-Braithwaite, S. 1996. *The English Comparable Corpus (ECC): A Resource and Methodology for the Empirical Study of Translation*, PhD Thesis, Manchester, UK, UMIST.

Partington, A., Duguid, A., and Taylor, Ch. (eds). 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Studies in Corpus Linguistics, volume 55. Amsterdam/Philadelphia: Benjamins.

Tiedemann, J. 2009. "News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces". In N. Nicolov, K. Bontcheva, G. Angelova and R

. Mitkov (eds.) Recent Advances in Natural Language Processing (vol V). Amsterdam/Philadelphia: Benjamins.

Wodak, R 2011. *The Discourse of Politics in Action. Politics as Usual*. London: Palgrave MacMillan.

Reshaping corporate image through translation: a corpus-assisted study

Sara Castagnoli
University of Macerata
sara.castagnoli@unimc.it

Elena Magistro
University of Bologna
elena.magistro2@unibo.it

Corporate Social Responsibility (CSR) reports are a form of voluntary corporate disclosure outlining a company's values and actions on social and environmental issues. Over the past few years, the business payoffs of CSR reporting have become evident: companies investing in CSR tend to benefit from enhanced reputation, increased trust and consumer loyalty; investors seem to be taking corporate sustainability into consideration to make investment decisions, as if a 'social contract' were governing the relationship between civil society and the corporate world (Fuoli 2012: 4; Bondi 2016). As such, CSR reporting becomes a discursive tool that seeks financial advantage by means of rational and moral legitimation: a company is legitimated to exist and operate – thus, to make profit – because of its social utility, promoting shared values and practically contributing to the common good (Bondi 2016; Van Leeuwen 2007; Cap 2005).

While not being precisely a form of advertising, CSR reports are intended to generate a positive evaluation and public image for the company, by combining classic promotional elements (e.g. self-praise, positive connotations) with elements aimed at bolstering and defending the legitimacy of corporate actions (Breeze 2012).

Although research on the discoursal and rhetorical features of CSR reports is extensive, with a minority of strictly linguistic studies, contrastive work on multilingual original reports is still limited, and even more so is the analysis of translated reports. The way translated CSR discourse fits in, and even impacts on, the target language and culture deserves greater attention: how companies shape their own image and establish relations with their stakeholders is closely connected to how they use pragmalinguistic elements, which reflect culturally-loaded sociopragmatic norms (Leech 1983).

This study is part of a larger corpus-assisted investigation of the effects of translation in sustainability reports, informed by a corpus-based contrastive analysis of original CSR reports in the languages examined, namely English and Italian. The corpus resource specifically designed for the study thus includes both a comparable component – i.e. original English vs. original Italian CSR reports – and a parallel component comprising source texts and translations in the two languages. The texts, produced between 2012 and 2017, were downloaded from an existing repository maintained by the Global Reporting Initiative, a non-governmental no-profit international organization providing (non-binding) standards for CSR reporting. The companies involved in the analysis are mostly large-sized and operate in a range of industries.

In a preliminary study (Castagnoli and Magistro, 2017) we analysed how companies shape their corporate *responsible persona* in original English and Italian CSR reports, and found that the degree of personalisation conveyed through first person plural deixis is significantly higher in the former. Translation-wise, however, we observed that personalisation patterns in English translations largely reproduce those found in Italian source texts, thus introducing in the English translated corpus features which may be perceived as 'alter' or simply not natural by the target reader (Venuti 1995; Magistro 2013).

The paper aims to extend and complement these findings through a comparative analysis of collocational patterns and lexical bundles for first person plural references in the different original and translated datasets with a view to identifying recurring strategies of self-representation in the reports, with a particular focus on specific pragmalinguistic resources (e.g. modal verbs) and lexical items (e.g. *sustainability*, *people*, *community*) that substantially contribute to constituting and shaping identities, perceptions and relations for companies and their stakeholders (Fairclough and Wodak 1997).

References

Breeze, R. 2012. "Legitimation in corporate discourse: Oil corporations after Deepwater Horizon". *Discourse & Society* 23(1): 3–18.

Bondi, M. 2016. "The future in reports: Prediction, commitment and legitimization in corporate social responsibility (CSR)". *Pragmatics and Society* 7(1): 57-81.

Cap, P. 2005. "Language and legitimization: Developments in the proximization model of political discourse analysis". *Lodz Papers in Pragmatics* 1: 7-36.

Castagnoli, S. and Magistro E. 2017. "Translating (Im)personalization in Corporate Discourse: A Corpus-Based Analysis of Corporate Social Responsibility Reports in English and Italian". Paper presented at the international conference *CLAVIER 2017 (Corpus and LAnguage Variation in English Research) - Representing and Redefining Specialised Knowledge*. Bari 'Aldo Moro' University, Bari, Italy, 20

November – 2 December 2017.

Fairclough, N. and Wodak, R. 1997. "Critical discourse analysis". In T.A. van Dijk (ed.) *Discourse as Social Interaction. Discourse Studies: A Multidisciplinary Introduction – Volume 2*. London: Sage. 258-284.

Fuoli, M. 2012. "Assessing social responsibility: A quantitative analysis of Appraisal in BP's and IKEA's social reports". *Discourse & Communication* 6(1): 55-81.

Leech, G. 1983. *Principles of Pragmatics*. London: Longman.

Magistro, E. 2013. "The challenges of 'translating' polite discourse for the EU multilingual community". *International Journal of Applied Linguistics* 23/1: 60-79.

van Leeuwen, T. 2007. "Legitimation in discourse and communication". *Discourse and Communication* 1(1): 91–112.

Venuti, L. 1995. *The Translator's Invisibility*. London and New York: Routledge.

The #ThingsITrustMoreThan CNN & FOX: A corpus-based exploration of the #thingsItrustmorethan movements

Alex Phillip Lyng Christiansen
Lancaster University & Aalborg University
a.christiansen1@lancaster.ac.uk

The present paper analyses the role of two Twitter 'hashtag' topics, #thingsItrustmorethanCNN and #thingsItrustmorethanFoxNews, seen as part of the continued (de)legitimization of news institutions as it exists on popular social media platforms, using the freeware tools *FireAnt* (Anthony & Hardaker 2017) and *AntConc* (Anthony 2017). Borrowing techniques from critical discourse analysis (CDA) and corpus linguistics (CL), as well as previous work utilizing Twitter in these fields (Zappavigna 2012; Hardaker & McGlashan 2016; McEnery, McGlashan, & Love 2015) the paper moves between varying aspects of discourse and practice, attempting, through CL methodology, to explain affiliative features of the two employed topics, and the language used to mark it.

Broadly speaking, the paper serves as a small part of a larger attempt at understanding the still growing tendency of anti-media sentiments. In a recent study by Poynter Media Studies, for example, Guess, Nyhan, & Reifler (2017) note a still growing divide in media trust on the political scale, with more than 60 percent of self-identifying pro-Trump voters showing "extreme[ly] negative attitudes" towards mainstream media, as well as a general belief that the media "keep[s] political leaders from doing their job." Support of particular news institutions seems symptomatic of the still more sharply defined division between republican and democrat voters, as conservative Americans favor Fox News and shun other institutions such as CNN, MSNBC and ABC, amongst others, and while liberal voters narrowly (16%) favour CNN, they overwhelmingly (87%) distrust Fox News (Mitchell et al. 2014: 14). These tendencies have bred numerous protests, ralies and movements, both on- and offline, and the #ThingsITrustMoreThan movement is one such protest; a recurring online topic whenever mistakes have been made by, especially, CNN, as well as a focus during earlier protests against e.g. Hillary Clinton, Barack Obama and Donald Trump. The employed corpus features a combined 15,000 tweets parted in two sub-corpora of 10,000 and 5,000 texts, both of which were collected on December 9th 2017, tangentially following the observed hashtags.

While the analysis highlights a number of expected features, such as similarities between the highlighted topics in the CNN corpus and those commonly found in Trump's Twitter discourse,

it also showed less initially obvious results. As such, while both sub-corpora show different aspects of anti-affiliative behaviour, often through derision of one or both of the observed institution, some data hints at users re-legitimizing institutions as well. Additionally, while the political left is commonly noted as having no central affiliation (Kasprak 2017), with e.g. only a small majority aligning themselves with a specific institution (Mitchell et al. 2014), the data shows clear consistency in practice and values, reflected in more consistently similar sentence structure and expressed sentiments. These observations bring into question the intentionality and community of the employed (anti-)media discourse, prompting a discussion of 'discourse communities', 'communities of practice' and 'communities of value'.

Apart from broadening the understanding of post-truth, anti-media sentiments, the paper serves as an example of issues and advantages in using Twitter as a corpus, without access to expensive enterprise data, providing a detailed explanation of collection and data-handling, especially commenting on newer developments in Twitter-functionality and issues with ascertaining the importance of retweets. For better or worse, it also serves as an example of ways in which inductive research methods may move research focuses over time and prompt a re-examination of both scope and topic.

References

Anthony, L. (2017). AntConc (Version 3.5.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Anthony, L. and Hardaker, C. (2017). FireAnt (Version 1.1.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software

Barthel, M., & Mitchell, A. (2017). *American's Attitudes About the News Media Deeply Divided Along Partisan Lines.* Pew. Pew Research Center. Retrieved from http://www.journalism.org

Guess, A., Nyhan, B., & Reifler, J. (2017). *"You're Fake News!".* St. Petersburg: The Poynter Institute for Media Studies.

Hardaker, C., & McGlashan, M. (2016). "Real men don't hate women": Twitter rape threats and group identity. *Journal of Pragmatics*, 80-93.

Kasprak, A. (2017, August 17). *Is there an 'Alt-Left'?* Retrieved from Snopes: http://www.snopes.com

McEnery, T., McGlashan, M., & Love, R. (2015). Press and social media reaction to ideologically inspired murder: the case of Lee Rigby. *Discourse & Communication*, 237-259.

Mitchell, A., Gottfried, J., Kiley, J., & Matsa, K. E. (2014). *Political Polarization & Media Habits.* Pew Media Studies. Retrieved from http://www.journalism.org

Zappavigna, M. (2012). *Discourse of Twitter and Social Media.* London: Continuum International Publishing Group.

Stylistic variation on the Donald Trump Twitter account

Isobelle Clarke
University of Birmingham
imc789@bham.ac.uk;

Jack Grieve
University of Birmingham
J.Grieve@bham.ac.uk

The 21st century has seen a rise in politicians using various forms of social media, especially Twitter. The Twitter account for the President of the United States (@realDonaldTrump) is arguably the most influential of these accounts in the world and an integral part of his campaign, election, and presidency. While the media tend to focus on the style and content of individual Tweets, there has yet to be a systematic stylistic analysis of all the Tweets posted on @realDonaldTrump. The goal of this research is therefore to describe the main dimensions of

stylistic variation found on the @realDonaldTrump Twitter account through a Multi-dimensional Analysis (MDA) (Biber, 1988).

To analyse stylistic variation on the @realDonaldTrump account, we downloaded all the Tweets from the Trump Twitter Archive (Brown, 2017), excluding re-Tweets, leaving 21,320 Tweets in our final corpus. We then tagged the Tweets for parts-of-speech information using the Gimpel et al. (2011) Twitter tagger and counted the occurrence of a wide range of grammatical features commonly analysed in MDA studies (Biber 1988), as well as a number of additional features associated with Twitter (e.g. hashtags, mentions), retaining the 65 features that occurred in at least 5% of the Tweets.

The standard approach to MDA is to subject the relative frequencies of a wide range of grammatical features in a corpus to a factor analysis to identify the most important dimensions of variation. Each of these dimensions is then interpreted functionally based on the individual linguistic features and the individual texts that are most strongly associated with them. The problem, however, with applying MDA to short texts like Tweets is that the relative frequencies of most grammatical forms cannot be measured reliably. Consequently, MDA analyses are generally limited to texts of at least several hundred words. To overcome this issue we applied a new form of MDA for the analysis of short texts (Clarke and Grieve, 2017). First, rather than computing the relative frequencies of grammatical forms, we simply recorded whether or not each of these forms occur in each of the Tweets. Second, rather than applying factor analysis, which is designed for analysing a set of continuous variables, we apply a multiple correspondence analysis, which is an analogous technique designed for analysing a set of categorical variables. Finally, we interpret the resulting dimensions in the standard way.

Overall, we found 5 main dimensions of stylistic variation, after controlling for text length. The first dimension contrasts Tweets with an overtly opinionated style to Tweets with a non-opinionated style. The second dimension contrasts Tweets focusing on predicting future events with Tweets concerning current issues. The third dimension contrasts Tweets that give advice with Tweets that do not. The fourth dimension contrasts Tweets with a promotional style with Tweets that have a less personal focus. And the fifth dimension contrasts Tweets with a critical style with Tweets that are more positive. Finally, we conclude our paper by discussing how our results help explain the influence of the Trump Twitter account.

References

Biber, D. (1988) *Variation across speech and writing*. Cambridge: Cambridge University Press

Brown, B. (2017) Trump Twitter Archive. Retrieved on 31 November 2017 from http://www.trumptwitterarchive.com

Clarke, I. and Grieve, J. (2017). "Dimensions of Abusive Language on Twitter". In *Proceedings of the First Workshop on Abusive Language Online*, pp. 1--10. Vancouver, Canada: ACL (August 4, 2017).

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J and Smith, N. A. (2011) Part-of-Speech tagging for Twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Short Papers* 19--24. Portland, Oregon: ACL (June 19-24, 2011).

'Unwanted gay thoughts keep shooting through my head': SO-OCD Sufferers' Discursive Construction of Ego-Dystonic Stance

Elvis Coimbra Gomes
Queen Mary University of London
e.coimbragomes@qmul.ac.uk

This poster will illustrate how corpus-assisted discourse analysis can widen our understanding on the often-misunderstood manifestations of sexual intrusive thoughts. 1-2% of the general population is affected by Obsessive-Compulsive Disorder (OCD) (APA 2013), from which up to 25% experience sexual intrusive thoughts (Grant et al 2006). The range of content of sexual intrusive thoughts 'is limited only by the imagination of the afflicted individual' (Gordon 2002: 344). People might obsessively doubt their sexual identity (Williams 2008; Goldberg 1984), gender identity (Safer and Safer 2016), or whether they are child molesters (O'Neil et al. 2005), among others. Although OCD is stereotypically associated with a compulsive urge to wash hands, the wider (clinical) population is unaware that OCD can hold on to such sexual themes. This results in sexual intrusive thoughts being very often misdiagnosed (Glazier et al. 2013). Indeed, although the diagnostic criteria to identify sexual intrusive thoughts are based on sufferers' ego-dystonic stance towards their unwanted thoughts, there seems to be a lack in the literature on how sufferers discursively express such stance. The current study aims to fill this important gap, since shedding light on these discursive strategies could help therapists' diagnostic practices.

With the raise of the Internet, more and more OCD sufferers found support in online forums (Stein 1997). Interestingly, it seems that the last decade has witnessed an increase of heterosexual people confessing to suffer from Sexual Orientation OCD (SO-OCD), where they anxiously believe to be turning gay. By using a self-compiled specialized corpus (ca. 227'000 tokens) of such online forum messages written by 299 male sufferers, I will apply corpus-assisted discourse analysis (Baker 2006) to focus on the keywords *thought(s), mind(s), head(s), brain(s)*. These findings will be compared to a smaller corpus (ca. 50'400 tokens) of messages written by 102 men who question their sexuality in a non-obsessive way.

Similar to Harvey's (2012) and Cassell's (1976) observations, initial findings illustrated that SO-OCD sufferers construct their *mind(s)*, *brain(s)*, and *head(s)* as containers intruded by their thoughts. This intrusive quality is discursively articulated by qualifying their thoughts with adjectives (e.g. *gay thoughts, unwanted thoughts, bad thoughts, irrational thoughts,* etc.); by specifying them with determiners (e.g. *a/the/that/those thoughts*); or by conveying them some uncontrollable agency (e.g. *these thoughts tell me that…* instead of *I thought that…*).

References

APA (American Psychiatric Association). 2013. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: Author.

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

Cassell, E. (1976). "Disease as an 'it': concepts of disease revealed by patients' presentation of symptoms". *Social Science and Medicine* 10(3), 143-146.

O'Neil, S.E., Cather, C., Fishel, A.K., & Kafka, M. 2005. "'Not Knowing If I Was a Pedophile …': Diagnostic Questions and Treatment Strategies in a Case of OCD". *Harvard Review of Psychiatry* 13(3), 186–196.

Glazier, K., Calixte, R.M., Rothschild, R., Pinto, A. 2013. "High Rates of OCD Symptom Misidentification by Mental Health Professionals". *Annals of Clinical Psychiatry* 25(3), 201-209.

Goldberg, R.L. 1984. "Heterosexual Panic". *The American Journal of Psychoanalysis* 44(2), 209-211.

Gordon, W.M. 2002. "Sexual Obsessions and OCD". *Sexual and Relationship Therapy* 17(4), 343-354.

Grant, J.E., Pinto, A., Gunnip, M., Mancebo, M.C., Eisen, J.L., Rasmussen, S.A. 2006. "Sexual Obsessions and Clinical Correlates in Adults with Obsessive-Compulsive Disorder". *Comprehensive*

Psychology 47, 325-239.

Harvey, K. 2012. Disclosures of depression: using corpus linguistics methods to interrogate young people's online health concerns. *International Journal of Corpus Linguistics* 17(3), 349-379.

Safer, D.L., Bullock, K.D., Safer, J.D. 2016. Obsessive-Compulsive Disorder Presenting as Gender Dysphoria/Gender Incongruence : a Case Report and Literature Review. *AACE Clinical Case Reports* 2(3), e268-e271.

Stein, D.J. 1997. Psychiatry on the Internet: survey of an OCD mailing list. *Psychiatric Bulletin* 21(2), 95–98.

Williams, M.T. 2008. Homosexual Anxiety: a Misunderstood Form of OCD. In L.V. Sebeki (ed.). *Leading-Edge Health Education Issues*. New York: Nova Biomedical Books. 195-205.

How do we talk about drought? The context of Britain as represented by national newspapers and oral histories

Carmen Dayrell
Lancaster University
c.dayrell@lancaster.ac.uk

Helen Baker
Lancaster University
h.baker1@lancaster.ac.uk

Rebecca Pearce
University of Exeter
R.Pearce@exeter.ac.uk

Tony McEnery
Lancaster University
a.mcenery@lancaster.ac.uk

Droughts are a global problem but no less evident in Britain. However, droughts are not necessarily caused by lack of rainfall alone; they can be worsened by a number of socio-economic factors such as the size of population and their rate of water consumption. Despite droughts being accepted as problematic, they are a creeping environmental phenomenon; as such the impacts manifest slowly and tend not to provoke immediate reactions, particularly when water provision is carefully regulated to protect users from the inconveniences associated with loss of supply.

In present-day Britain, dialogues around droughts are closely associated with the management of the commoditisation of natural resources, following the privatisation of water supplies across England. They therefore relate directly to power relations between water utilities, water regulators, and private and public customers. In order to improve our ability to manage instances of future drought, we would benefit from a greater understanding of the public discourses about issues concerning appeals for conservative use of water when supplies are limited.

This paper will examine public discourses around droughts and water scarcity events in the UK from the perspectives of national newspapers and oral histories. By oral histories we refer to people's memories of past drought events. These are three separate bodies of evidence, collected independently and guided by different methodological approaches. The novelty of this research is to combine two different methodological approaches – corpus linguistics and memory retrieval – to explore public discourses around historic droughts. The analysis covers the time period of 1976 to date. The 1976 drought is considered the most severe in the climate history of 20th-century Britain, with sustained impacts across the whole of the UK. Memories of this drought are retained by a large portion of the British population. We are also interested in subsequent localised and less intense drought events, such as those of 1995 and 2006, which featured nationally in public discourses and are evident in private recollections.

We have used two separate corpora of national newspapers. For contemporary newspapers we have considered texts published by all major British broadsheet and tabloid newspapers, going back in time as far as data was available in the news aggregator service (*Lexis Nexis*). The corpus covers the time period of 01/01/1990 to 31/12/2014, comprising 4,986 texts (3.8 million words) from broadsheet papers and 2,384 texts (1.1 million words) from tabloids. For the period between 1976 and 1999, we used *The Times* corpus which contains a full-text copy of the OCR texts underlying *The Times Online Archive*. For the analysis, we have utilised GIS (Geographic Information Systems) methods to distinguish between instances referring to drought events in the UK specifically from those occurring outside the UK. Discourse patterns in the national newspaper data were examined through the analysis of the collocations networks of the word *drought** and interpreted by close reading of the concordance lines.

Oral history recordings were sought from donors in locations identified from hydro meteorological records as areas that had been severely affected by droughts, and also from professions that were considered likely to have been directly affected by droughts. From 69 detailed personal recollections of the impacts of droughts on the lives and livelihoods of individuals, a search of local newspapers from the locations and periods of discussion yielded 579 corresponding articles, letters, water management reports, and water saving appeal advertisements.

This paper will discuss the extent to which personal memories of drought events are compatible with reporting in national newspapers of the time and how effectively the presentation and narratives surrounding drought in news texts captured perceptions and the experiences of droughts in everyday lives. While the national press presented a consistent narrative across time, providing a general account of the impact of droughts across Britain and measures in force to ensure maintenance of supplies, local newspapers and oral histories provided descriptive accounts of droughts in specific places, thus offering details of the experiences of those who have been primarily affected.

Metaphors and narratives of climate change across genres and discourse communities: A corpus-based comparison

Alice Deignan
University of Leeds
A.H.Deignan@education.leeds.ac.uk

Elena Semino
Lancaster University
e.semino@lancaster.ac.uk

Climate change is one of the world's most urgent issues. Young people, in particular, are likely to be affected in their lifetimes, and will also influence future developments through their lifestyle choices and decisions as citizens. As with other scientific topics, however, knowledge about climate change is mediated through discourse. In this paper, we are particularly concerned with the relationship between representations of climate change for and by members of different discourse communities: scientists, school teachers and secondary school students. We therefore present the results of a comparative analysis of three corpora, consisting of: (1) Academic articles about climate science (approximately 500,000 words); (2) Educational materials for secondary school students in the UK (approximately 250,000 words); (3) Interviews with secondary school students in the North of England (approximately 90,000 words). The three corpora were analysed using: word frequency lists; key words and key semantic domains; and

selected lexical and semantic concordances. We will focus particularly on the findings that relate to how the students' use of metaphors and narratives for climate change contrasts with the metaphors and narratives identified in the other two corpora.

Both metaphors and narratives are well known to be important tools for communication and sense-making, including particularly in the domains of science and education. We show two main types of differences in metaphor use between the student interviews and the other two corpora, i.e.: metaphors that are only used by the students (e.g. a 'band' around the earth that gets 'tighter and tighter'), and metaphors that are shared with one or both other corpora, but are used differently by the students. The latter applies particularly to the 'greenhouse' metaphor for global warming. In the academic corpus, 'greenhouse' is a technical term which only occurs as a pre-modifier in a limited range of noun phrases (e.g. 'greenhouse gases'). In the student interviews, it is a live metaphor that encourages the creative use of everyday knowledge about greenhouses and glass structures to talk and reason about climate change (e.g. 'My mum has a greenhouse so I kind of like refer back to that'). While the creative exploitation of the 'greenhouse' metaphor showed active engagement with scientific language and ideas, it also appeared to lead to some misconceptions, for instance, that greenhouse gases form a thin, hard shell around the earth analogous to glass. Similarly, the narratives that the students produced to describe the future effects of climate change seemed to both exaggerate and oversimplify the complex science of climate change. These narratives drew from semantic domains of death and destruction and ended with descriptions of events such as seas boiling and animals melting. We suggest that the extreme and apocalyptic nature of the students' narratives is more reminiscent of the plots of popular representations of climate change in blockbuster movies than of the scenarios presented in teaching materials and scientific articles.

We finish with some reflections on the potential implications of these metaphors and narratives for people's views and actions in relation to climate change. On the one hand, it is predictable and perhaps inevitable that understandings of complex scientific phenomena such as climate change will differ between members of 'lay' and 'expert' discourse communities. On the other hand, our findings suggest that some of the (mis)understandings identified in our student corpus might have negative consequences. For example, they could lead young people to underestimate reports that average global temperature have risen by one or two degrees centigrade. We therefore suggest that there is a need for better understandings of the nature of metaphors and narratives in scientific reasoning in the school science classroom, and make some recommendations for the future design of teaching materials.

The patient is always right: A corpus-based analysis of the term 'customer service' in online patient feedback

Craig Evans
Lancaster University
c.evans4@lancaster.ac.uk

The notion of patients being customers is a subject of ongoing interest in health research (Hudak, McKeever, & Wright, 2003; Powers et al., 2013). In the UK context, where commercialisation can be viewed as a precursor to NHS privatisation (Brookes & Harvey, 2016), the patient-customer is thriving; at least, that is, in the language use of patients when

describing their healthcare experiences. In a corpus of online patient feedback from the review website NHS Choices, 'customer' is one of the strongest collocates of the keyword 'service', producing the term 'customer service' 95% of the time they collocate. It is perhaps not surprising that 'customer service' should be salient on a review website, a communicative situation usually associated with commercial practice. However, from a critical discourse analytic perspective, this does not automatically mean the discourse is being colonised by a culture of enterprise (Cameron, 2000; Fairclough, 1993). What an individual term tells us about discourse depends on patterns of its use in context. This study thus investigates the term 'customer service' with the aim of answering the question: what is the significance of this lexical item when used in healthcare discourse?

This paper reports on a collocation and concordance analysis of 'customer service' in a 29-million-word corpus of online patient comments from NHS Choices, 2013–2015. Using the concept of discourse prosody (Baker, 2006), an account of healthcare discourse is provided by identifying collocation patterns. These are then analysed at several levels of context including: the surrounding text, the interactional circumstances (i.e. patients posting comments in a public space to which NHS staff often respond), and the wider socio-political situation of present-day UK healthcare.

The collocation analysis highlighted reference to several categories, most notably 'training', 'negative evaluation', 'positive evaluation', and 'measures and amounts'. In this study, collocates referencing 'training' often entail a call to action, namely staff training, and are usually found in comments where staff impoliteness is being criticised. Such comments suggest the neoliberal view that enhancing individual expertise is the solution to poor 'customer service'. This backgrounds structural reasons for perceived poor customer service, such as staff abruptness being due to ever-increasing efficiency demands when spending cuts lead to limited resources (Mason, 2014). Collocates that reference 'negative evaluation' and 'measures and amounts' indicate a dramatisation of patients' experience, as in the intensifier 'absolutely' (e.g. 'absolutely no') and emotive adjectives like 'shocking' (e.g. 'horrific shocking customer service'). Such collocates serve a persuasive purpose and perhaps reflect differences in staff–patient power dynamics in an online space, where patients have a public platform to express themselves, compared with a physical space, where patients are subject to the procedural constraints of healthcare practices.

Analysing collocates of 'customer service' has revealed that the commercial values associated with neoliberalism are inherent in patients' language use. There is some evidence – in replies to comments – of these values coming into practice, such as when staff report on acting on calls for training, although this primarily concerns soft skills training. Collocates that suggest patients might be empowered when reviewing 'customer service' raise questions about how empowered they really are as customers of a service that still mostly exists outside the market.

References

Baker, P. (2006). *Using corpora in discourse analysis*. London: Continuum.

Brookes, G., & Harvey, K. (2016). Opening up the NHS to market: Using multimodal critical discourse analysis to examine the ongoing commercialisation of health care. *Journal of Language and Politics, 15*(3), 288–303.

Cameron, D. (2000). *Good to talk?* London: Sage.

Fairclough, N. (1993). Critical discourse analysis and the marketisation of public discourse: The universities. *Discourse & Society, 4*(2), 133–168.

Hudak, P., Mckeever, P., & Wright, J. (2003). The metaphor of patients as customers: Implications for

measuring satisfaction. *Journal of Clinical Epidemiology, 56*(2), 103–108.

Mason, R. (2014, February 23). GP services under pressure due to funding cuts, says Royal College of GPs. *The Guardian*. Retrieved from: www.theguardian.com/society/2014/feb/23/gp-services-patients-funding-nhs-hospitals.

Powers, B., Navathe, A., Aung, K., & Jain, S. (2013). Patients as customers: Applying service industry lessons to health care. *Healthcare, 1*(3–4), 59–60.

A quantitative evaluation of keyword measures for corpus-based discourse analysis

Stefan Evert	**Natalie Dykes**	**Joachim Peters**
University of Erlangen-Nuremberg	University of Erlangen-Nuremberg	University of Erlangen-Nuremberg
stefan.evert@fau.de	natalie.mary.dykes@fau.de	joachim.peters@fau.de

We evaluated the automatic identification of keywords in our target corpus of German press texts on multi-resistant pathogens; a recurring and highly controversial topic in public discourse (1.3M tokens). In corpus-based discourse analysis, keywords have been described as "a quick and simple 'way in'" to corpus comparison (Baker et al. 2013), while the details on how to best 1) calculate and 2) categorise them lacks consensus. There have been multiple attempts to answer question 1) (cf. Kilgarriff 2001, Paquot & Bestgen 2009, Lijffijt et al. 2016). However, these studies focus mostly on either the number of keywords or the mathematical adequacy, but do not evaluate against an authentic use-case specifically tailored to discourse analysis. Our approach draws on a set of previously determined qualitative linguistic categories and evaluates statistically generated keyword lists against them. The keywords were calculated using three measures:

- log-likelihood (G^2) (Dunning 1993)
- log ratio (LR), an intuitive measure implemented in the CQPweb corpus analysis software (Hardie 2012)
- LR_{conf}, a conservative estimate for the log ratio coefficient using the lower bound of a 99% confidence interval with Bonferroni correction and two different German national broadsheet newspapers as reference corpora (RC):
- years 2011–2014 of *Süddeutsche Zeitung* (SZ), a left-leaning daily newspaper (290M tokens)
- years 2011–2014 of *Frankfurter Allgemeine Zeitung* (FAZ), a right-leaning daily newspaper (150M tokens)

All corpora were POS-tagged and lemmatized using the state-of-the-art morphological analyzer SMOR (Schmid et al. 2004).

Applying a frequency threshold of $f \geq 5$ in the RC, we obtained the 200 top-scoring keywords for each combination of keyness measure and RC. The various measures show substantial differences (Figure 1). The results suggest that LR_{conf} is intermediate between the other two measures, but it is closer to LR than G^2. The choice of the RC has more influence than expected – according to the literature, the keywords should be relatively stable for large RC (Scott & Tribble 2006). While the top-200 lists against SZ vs. FAZ have an overlap of 89.0% for G^2, the overlap is only 78.0% for LR_{conf} and 58.5% for LR. Especially for LR, which tends to select

lower-frequency keywords, the differences are partly due to the frequency threshold on the RC; but even for keywords occurring $f \geq 5$ times in both RC, the ranking differences are bigger than for G^2. A crucial difference between the keyness measures is that G^2 prefers keywords with high frequency in the target corpus, while LR is biased towards low-frequency keywords (Hardie 2014) . Figure 2 confirms this expectation, showing LR_{conf} as a compromise between the two extremes.

Our evaluation is based on a manual classification of the keywords into 33 categories previously identified in a detailed qualitative analysis (Peters 2017). A keyword is considered a true positive (TP) if it can be clearly assigned to a category or expresses positive or negative sentiment, otherwise it is considered a false positive (FP). All six lists were pooled for the manual annotation, resulting in a total of 455 distinct lemmas.

Figure 3 shows the precision achieved for each keyness measure and each of the two RC. Overall, keyword identification works fairly well with precision ranging from 60.5% to 66.0%. Differences between the measures and RC are small, with LR and LR_{conf} slightly better than G^2. None of the differences are significant according to a McNemar-style test (Evert 2004).

In corpus-based discourse analysis, precision is only of secondary concern: manually discarding many FPs may be tedious, but does not lower the quality of the final analysis if sufficient support can be found for all relevant categories. The main aspect of our evaluation is thus on the recall of top-200 keyword lists, i.e. how much support they give to each category. The plot below shows the number of keywords supporting each category. Results are shown for the larger SZ RC; those for FAZ look very similar.

A key observation is that some categories have substantial support from automatically extracted keywords, whereas others – notably metaphors – can hardly be inferred from the corpus-based analysis. There are moderate differences between the keyness measures: G2 identifies many evaluative terms, whereas LR finds more keywords related to pathogens, hospitals and scientific background. However, a corpus-based analysis using either measure would likely come up with the same categories (and also miss the same set, especially metaphors). Surprisingly, LRconf does not appear to be simply a compromise between G2 and LR; in some categories, it is even more extreme than LR.

References

Baker, P., Gabrielatos, C., and McEnery, T. 2013. *Discourse analysis and media attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Dunning, T. 1993. "Accurate methods for the statistics of surprise and coincidence". *Computational linguistics*, 19(1), 61–74.

Evert, S. 2004. "Significance tests for the evaluation of ranking methods". *Proceedings of the 20th International Conference on Computational Linguistics*. Geneva, Switzerland: COLING 2004. 945–951

Hardie, A. 2012. "CQPweb – combining power, flexibility and usability in a corpus analysis tool". *International Journal of Corpus Linguistics*, 17(3), 380–409.

Hardie, A. 2014. "A single statistical technique for keywords, lockwords, and collocations". Internal CASS working paper no. 1, unpublished.

Kilgarriff, A. 2001. "Comparing corpora". *International Journal of Corpus Linguistics*, 6(1), 97–133.

Lijffijt, J., Nevalainen, T., Säily, T., Papapetrou, P., Puolamäki, K., and Mannila, H. 2016. "Significance testing of word frequencies in corpora". *Digital Scholarship in the Humanities*, 31(2), 374–397.

Paquot, M., and Bestgen, Y. (2009). "Distinctive words in academic writing: a comparison of three

statistical tests for keyword extraction". In A. Jucker,, D. Schreier, and M. Hundt (Eds.) *Corpora: Pragmatics and Discourse. Papers from the 29th International Conference on English Language Research on Computerized Corpora*. Amsterdam: Rodpoi. 247–269.

Peters, J. 2017. Den Feind beschreiben. Multiresistente Erreger im deutschen Pressediskurs. Eine diskurslinguistische Untersuchung der Jahre 1994–2015 (master's thesis). Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, unpublished.

Schmid, H.; Fitschen, A.; Heid, U. 2004. "SMOR: A German computational morphology covering derivation, composition, and inflection". *Proceedings of the 4th International Conference on Language Resources and Evaluation*. Lisbon, Portugal: LREC 2004. 1263–1266

Scott, M., and Tribble, C. 2006. *Textual patterns – Key words and corpus analysis in language education*., Philadelphia: John Benjamins.
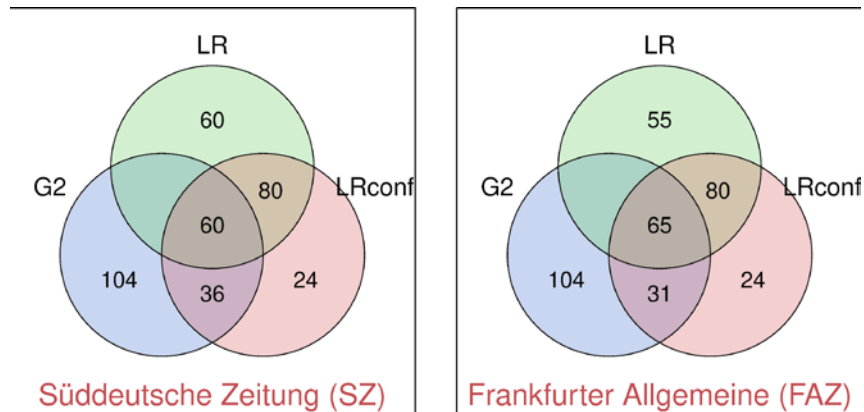

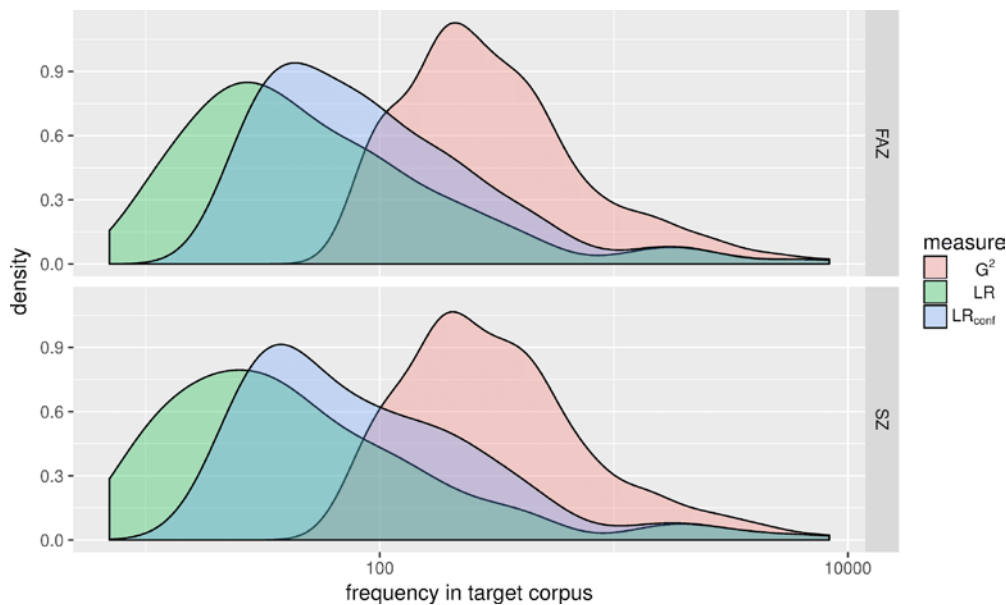*Figure 1: Overlap between top-200 keyword lists for each RC.*

Figure 2: Distribution of target corpus frequency in top-200 lists according to different keyness measures.
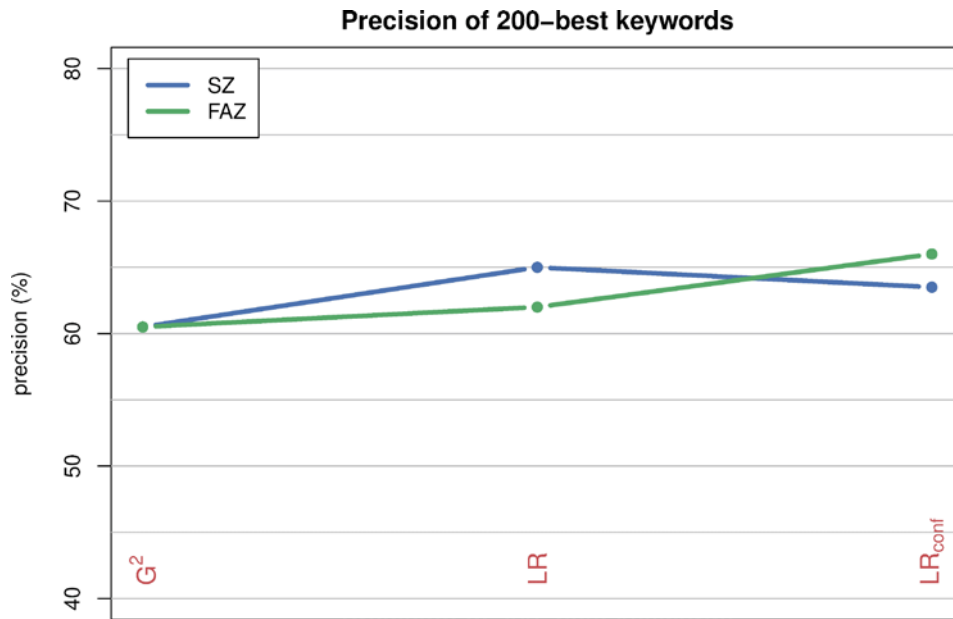
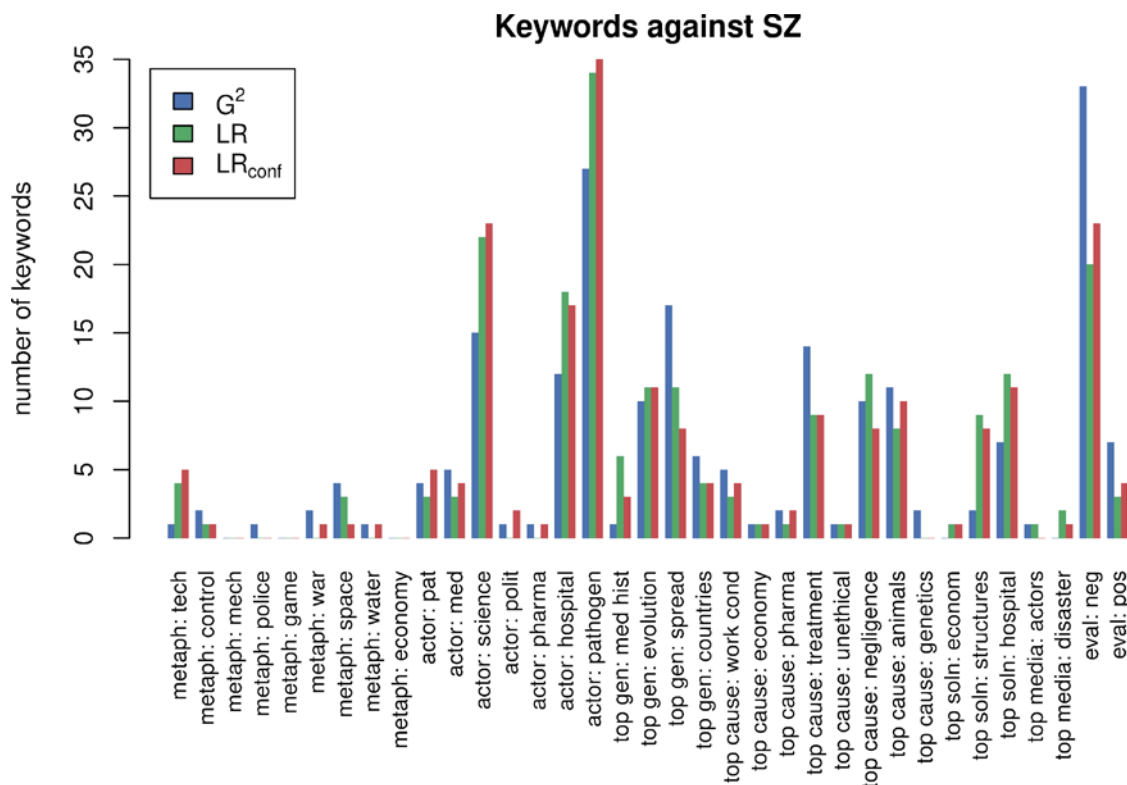*Figure 3: Precision of the top-200 keyword lists.*

Figure 4: Number of keywords supporting each category among top-200 lists for three different keyness measures.

"One hundred times a better person": A corpus based critical discourse analysis of the International Baccalaureate (IB) in Canadian newspapers

Saira Fitzgerald
Carleton University
saira_fitzgerald@carleton.ca

This study combines corpus linguistics and the discourse historical approach (DHA) branch of critical discourse analysis (CDA) (e.g., Baker et al., 2008; Potts, 2015) to examine how the International Baccalaureate (IB) is discursively constructed in Canada. The IB is a series of educational programs for students aged 3 to 19 offered in schools around the world. When it was originally created 50 years ago, the IB was a niche product for the children of diplomats studying in private international schools who needed a portable and recognized curriculum as they moved from country to country. The picture of the IB today is very different. Instead of international private schools as its milieu, the IB has grown in popularity in publicly funded schools in Australia, the UK, the US and Canada, with the highest number (2265 [45%]) of schools located in North America. This move into publicly funded education becomes problematic as a two-tier private-in-public school system is created, privileging some and disadvantaging others.

To gain insight into the dominant perceptions of the IB in Canada, a specialized corpus of newspapers was created as an unsolicited window into public opinion (Mautner, 2008). Articles were obtained through two online news databases, Canada Newsstand Complete and LexisNexis Academic, using the search terms *international* AND *baccalaureate, ib* AND *diploma* OR *student.* All Canadian newspapers (national, regional, local, and community (paid or free)) available at the time of the search were included to ensure maximum representation of public attitudes and perceptions (McEnery & Hardie, 2012; Seale, 2003). To allow for the broadest possibility of voices/perspectives, i.e., not just powerful elite sources but parents, students, and other members of the public, all types of newspaper texts were included in the corpus, such as letters to the editor, opinion pieces, obituaries, news and sports. The aim was not to examine the newspapers themselves but rather to gain a window into the way the IB is viewed by different social domains as represented by the different types of texts.

The corpus tools used for the analysis were AntConc (Anthony, 2014), COCA (Davis, 2015) and Sketch Engine (Kilgarriff et al., 2014), chosen not only because they provide different functionalities, but also as a further means of triangulation. Different tools offer different windows into the data which allows for a more complete picture to emerge (Taylor, 2010).

Results from collocation and concordance analysis revealed a positive discourse prosody (Stubbs, 2001) with underlying ideas of quality and morality, and also suggested that the positive view of the IB tends to create a negative view of things non-IB (programs, students, schools). Since one is constructed as "better", there is an implied comparison that seems to go unnoticed. Patterns of language use were analyzed to make visible values and assumptions that discursively construct the IB as superior. The linguistic patterns and strategies identified bear a striking resemblance to discourses of discrimination and difference (KhosraviNik, 2010; Reisigl & Wodak, 2001). By highlighting how groups may be excluded or disadvantaged, this study hopes to problematize the seemingly uncritical acceptance of IB programs in Canadian public

schools.

References

Anthony, L. (2014). *AntConc* (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273-306.

Davis, M. (2015). *The Corpus of Contemporary American English (COCA): 520 million words, 1990-2015.* Available at http://corpus.byu.edu/coca/

KhosraviNik, M. (2010). The representation of refugees, asylum seekers and immigrants in British newspapers: A critical discourse analysis. *Journal of Language and Politics, 9*(1), 1-28.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The sketch engine: Ten years on. *Lexicography, 1,* 7-36.

Mautner, G. (2008). Analyzing newspapers, magazines and other print media. In R. Wodak & M. Krzyżanowski (Eds.), *Qualitative discourse analysis in the social sciences* (pp. 30-53). New York, NY: Palgrave Macmillan.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics.* Cambridge: Cambridge University Press.

Potts, A. (2015). Filtering the flood: Semantic tagging as a method of identifying salient discourse topics in a large corpus of Hurricane Katrina reportage. In P. Baker & T. McEnery (Eds.), *Corpora and discourse studies: Integrating discourse and corpora* (pp. 285-304). London, UK: Palgrave MacMillan.

Reisigl, M., & Wodak, R. (2001). *Discourse and discrimination: Rhetorics of racism and antisemitism.* London, UK: Routledge.

Seale, C. (2003). Methodology versus scholarship? Overcoming the divide in analysing identity narratives of people with cancer. *Journal of Language and Politics, 2*(2), 289-309.

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics.* Oxford: Blackwell.

Taylor, C. (2010). Science in the news: A diachronic perspective. *Corpora, 5*(2), 221-250.

Negotiating stance in abortion discourse: a corpus-based analysis

Kristen Fleckenstein
The University of Texas at Arlington
kristen.fleckenstein@mavs.uta.edu

Currently, there are competing theories of stance-taking as to whether individuals privilege expression of personal opinion (Myers 2010) or constructing group membership (Paterson et al. 2016) in discourse; it remains unclear whether this discrepancy results from language register or the discourse topic. The present study examines whether register variation influences which type of stance a writer privileges by studying written discourse on abortion.

Two corpora were collected for this study: one from editorials in U.S. media sources (53,338 words), and one of informal discourse from the social media website Reddit (84,375 words). Within these corpora, three indicators of stance-taking were examined: modals of necessity/obligation, agentless passives, and rhetorical questions.

In the editorial corpus, interpersonally polite modals like *should* and *have to* are collectively 59.8% of modals used. This contrasts with more forceful modals like *must* and *need to*, which represent only 24.6% of total modals. The Reddit corpus shows a similar pattern, with *should* and *have to* comprising 70.3% of modals compared to 14.9% for *must* and *need to*. The use of softer modals over more forceful ones represent instances of face-saving politeness (Gales

2010) towards the reader.

    (1) a. We should all agree, [...], that abortion is not a solution to the host of systemic injustices driving

          poverty (Editorial corpus).

        b. It's MY body, YOU have to justify yourself to ME [...] (Reddit corpus)

Writers in both registers make use of agentless passives, which allow the writer to foreground their view on the patient of the sentence while maintaining the appearance of neutrality (Baratta 2009).

    (2) a. Bad clinics [...] should be closed immediately (Editorial corpus).

        b. If someone didn't consent to sex, they shouldn't be forced to consent to the consequences

          (Reddit corpus).

Rhetorical questions demonstrate register variation between the corpora, as they were only present in informal discourse; the Reddit corpus features several instances of rhetorical questions, where the questions function as reported thought (Kim 2014) to position themselves as taking an adversarial or challenging stance toward other writers within the discourse.

    (3) Why should fetuses get a right that no born person has? (Reddit corpus)

The use of different stance-taking devices reveals an underlying tension in the writers' stance in informal contexts. The writers make use of both adversarial stances that signal the expression of their personal opinion as well as stance-taking devices that mitigate face-threatening acts, which would seem to emphasize the importance of group membership. Formal discourse shows a stronger emphasis on group membership through politeness forms and passives. The register of language use seems to influence which stances writers privilege.

References

Baratta, A. M. 2009. Revealing stance through passive voice. *Journal of Pragmatics* 41. 1406-1421.

Gales, T. A. 2010. *Ideologies of violence: A corpus and discourse analytic approach to stance in threatening communications*. Doctoral dissertation, University of California, Davis.

Kim, M.S. 2014. Reported thought as a stance-taking device in Korean conversation. *Discourse Processes* 51(3). 230-263.

Myers, G. 2010. Stance-taking and public discussion in blogs. *Critical Discourse Studies* 7(4). 263-275.

Paterson, L., L. Coffey-Glover, & D. Peplow. Negotiating stance within discourses of class: Reactions to *Benefits Street*. *Discourse & Society* 27(2). 195-214.

A *terrible crime* for futile motives and/or unconditional love: a corpus-investigation of 'femminicidio' in Italy

Federica Formato
Edge Hill University
Federicaformato.ac@gmail.com

The increasing number of men killing women with whom they have been in a relationship has alerted Italy to consider how gender is perceived and performed within a fixed understanding of *the* private sphere. A historical legacy which started with the Fascist Era is part of how women and men are seen operating in the private (and in the public) as well as how personal

relationships between them are conceived. Abrogated in the early 80s, Italy promulgated a law in 30s to legislate the so-called 'honour killing' which would diminish the punishment of a man killing his wife or his female relatives if the murder was "determinato dall'offesa recata all'onor suo" (determined by the offence to his honour, Codice Rocco, 1930, art.587).

With, arguably, part of the country maintaining this understanding, the Italian term 'femminicidio' seems to be now accepted as the only one that can describe the crime for which men deliberately murder women. In this crime, women are seen as powerless and as the weakest within a patriarchal culture. Other terms are used for contexts outside Italy, e.g. *domestic violence* or *intimate murders* in UK. The discussion on terminology embodies the problematization of why men murder women they have been in relationships with; in fact, the legal term *uxoricide* seems to disregard the symbolism of gender of the criminal event while basing its understanding on 'biological' gender (Dobash and Dobash, 1998). From 2013 – when a law passed in the parliament to strengthen the punishment for those who commit 'femminicidio' – more than 500 women have been killed by former or on-going partners. Sociologists as well as criminologists have discussed forms of gendered violence as belonging to a symbolic gendered belief system which is "crucial in understanding and deconstructing what happened and why" (Monckton-Smith, 2012: 2). Based on previous academic research on domestic abuse in other countries (Monckton-Smith, 2012; Tabbert, 2016), I investigate a news dataset to illustrate subtle and evident discourses around 'femminicidio'.

I use a CADS (Corpus assisted discourse studies, Partington, Duguid and Taylor, 2013) approach to examine the dataset as a *discourse type*, namely a 4-year opportunistic corpus of Italian newspaper articles on the topic of 'femminicidio' for the 2013-2016 period (total number of words: 131,033, total number of articles: 331) compiled through the database Nexis. This methodological and theoretical approach also allows for contextualization of the dataset (and the rationale for this research) within sources of information which are *outside* the corpus. Specifically, this study is also informed by the *Manifesto di Venezia*, a document signed by female and male journalists on 25 November 2017. In this manifesto, one can find guidelines on how to fairly narrate violence against women and femminicidio. Before it made its appearance, other activists – belonging to the group *Non Una di Meno* - spoke out in relation to a biased narration of violence, one which seemed to prefer the male offender's point of view while blaming the victim for not adhering to a fix prototype of (passive and un-promiscuous) womanhood.

Starting from these sources, I performed keyword analysis (single word and multi-word) comparing the news corpus against ItTenTen16 as a reference corpus of Italian news (with SketchEngine) in order to investigate what I define as 'forensic narratives' in detail. To explore these, I built a taxonomy arising from the keywords that shed lights on the gendered private space as well as *the* gendered groups of male offenders and female victims. Furthermore, I examine changes and similarities in the use of specific terms across the 4 years to investigate whether the news have adapted their style to a fairer narration of this criminal phenomenon.

To conclude, this study contributes to existing research in sociology and criminology, providing solid and consistent linguistic results, while also addressing the gap in research on the language – Italian – (and) on a specific criminal phenomenon.

References

Codice Penale. 1930. *Omicidio e lesione personale a causa di onore.* Art. 587, available at https://www.brocardi.it/codice-penale/libro-secondo/titolo-xii/capo-i/art587.html. Accessed on 28/12/2017.

Dobash, R.E. and Dobash, R.P. eds., 1998. *Rethinking violence against women.* London: Sage Publications.

Monckton-Smith, J. 2012. *Murder, gender and the media: Narratives of dangerous love.* Berlin: Springer.

Partington, A., Duguid, A., & Taylor, C. 2013. *Patterns and meanings in discourse: theory and practice in corpus-assisted discourse studies* (CADS). Amsterdam:John Benjamins Publishing.

Tabbert, U. 2016. *Language and Crime: Constructing Offenders and Victims in Newspaper Reports.* Berlin: Springer.

'Lesbians in Modern Day Britain: A Corpus-Based Study of The Discourses Surrounding the Lemma LESBIAN*

Frazer Heritage
Lancaster University
f.heritage@lancaster.ac.uk

This paper explores how lesbians are discussed in both informal spoken British English and the written British media. It draws on two data sources: the newly compiled spoken BNC2014 (Love et al., 2017; Love, 2018) and a specific corpus of 1.2 million words of data from newspapers written between January 2017 and December 2017 inclusive. I argue that the discourses produced and reproduced in both corpora are ways of sustaining and maintaining problematic ideological stances towards groups marginalised due to their sexual identity (see Baker, 2014; Gupta, 2016). The present study acts as a follow-up paper to Motschenbacher's (forthcoming) research into adjectival collocates of sexual identity within the Corpus of Contemporary American English (COCA). Where this paper and Motschenbacher's research differ is that I specifically elect to draw on spoken British English instead of American English. I argue that the current discourses surrounding lesbians found within the both the spoken BNC 2014 and the specific written corpus remain relatively problematic and archaic. I discuss two prominent discourses in spoken British English, two prominent discourses in written British English, and one discourse which overlaps across both corpora. I argue that the discourses in both corpora are damaging to the way lesbians are viewed in general society and that the discourses tend to mask an underlying homophobic ideology.

Discourses surrounding the lemma LESBIAN* in spoken language tend to suggest that that lesbianism is a choice and that lesbians who preform masculinity are 'undesirable'. In this paper, I give specific examples of these discourses and explore how they convey this underlying homophobic ideology towards lesbians. Within the written corpus, I employed keyword analysis in addition to concordance line analysis, which revealed that lesbians are seen with regards to legal policy and that they are sexually charged beings. I also employ concordance line analysis, which revealed discourses that positioned lesbians as sexual threats to children and as inauthentic compared to heterosexual identities. One issues which appears in both spoken and written modes is the positioning lesbians within the context of 'gay' rights.

References

Baker, P. 2014. *Using corpora to analyze gender.* London: Bloomsbury.

Gupta, K. 2016. *Representation of the British suffrage movement.* London: Bloomsbury.

Love, R., Dembry, C., Hardie, A., Brezina, V., & Mcenery, T. (2017). The Spoken BNC2014. Compiling and Analysing the Spoken British National Corpus 2014. *International Journal of Corpus Linguistics*, 22(3), 319-344.

Love, R. (2018). *The Spoken British National Corpus 2014: design, compilation and analysis.* PhD thesis. Lancaster University

Motschenbacher, H. (forthcoming). Corpus linguistics in language and sexuality studies: Taking stock and future directions. *Journal of Language and Sexuality,* 7 (2).

The Language of Legitimacy: A Computer-Assisted Study of the Discursive Legitimation of Global Courts

Zuzanna Godzimirska
University of Copenhagen
zuzanna.godzimirska@jur.ku.dk

International courts need legitimacy in order to function effectively, to ensure compliance with their decisions, and ultimately to survive. Given that existing scholarship has dedicated significant attention to the judicial means by which international courts gain and maintain legitimacy, this paper focuses its attention on the discursive level of legitimacy, analyzing the discursive interplay between international courts and nation states.

Drawing on two corpora consisting of judgments of two global courts – the International Court of Justice and the World Trade Organization's Appellate Body – and government statements on them in the period 1996-2016, the paper explores how international courts' claims to authority are constructed, deconstructed and reconstructed in discourse in their interface with their primary constituents, states.

The paper employs a combination of automated methods of text classification with corpus approaches to discourses analysis. Employing supervised learning methods, it first classifies paragraphs in the courts' judgments into two categories: those containing legitimation discourses and those in which such discourses are absent. This is followed by a classification of government statements into three categories: those that 1) delegitimate or 2) legitimate the courts, or 3) those in which legitimation discourses on the courts are absent. Units within both corpora in which (de)legitimation discourses are absent, are eliminated from the subsequent analysis. Units within court decisions and government statements that do contain (de)legitimation discourses are, on the other hand, brought together into different sub-corpora and compared to examine the differences in the linguistic resources that are used on and by international courts to legitimate these institutions. In operationalizing international court legitimation, the paper draws on Theo van Leeuwen's framework for analyzing the way discourses construct legitimation.

International courts have a number of sources of law and means of interpretation at their disposal to justify specific interpretation and findings. For example, in deciding a case, a court may refer to a provision text, case law of other international courts, or simply, rely on its own authority when pronouncing what the legal rule is in a given area. Through key word analysis, I will explore the differences in the way the two courts justify their exercise of authority in their judgments. Does one refer more to provision texts, or other courts' case law? Those one rely more on its own inherent authority when deciding a case? Do these patterns change when the courts' authority is challenged by their primary constituents, member states? Building on this

key word analysis, collocation analysis is then employed to examine, in more depth, the specific linguistic resources used to de- and legitimate the courts.

The aim of the paper is two-fold. First, it aims at increasing our understanding of how international courts claim or reclaim legitimacy, in the discursive interaction with their primary constituents, states. Second, it aims at adding a brick to the existing, but yet modest bridge between the disciplines of law and linguistics, and shedding new light on how corpus linguistic methods can assist in mapping the discursive process of legitimation of judicial institutions.

Mapping different discursive formation: how do populist parties depict Europe?

Juan Roch González
Freie Universität Berlin
juan.roch.gonzalez@fu-berlin.de

The study of populism and populist discourse is on the rise in political science. Commentators and political scientists tend to analyse different and sometimes antagonistic political formations under the conceptual insights provided by the study of populist phenomena. In response to this diversity, exclusionary and inclusionary forms of populism appear consolidated as subtypes within the literature on populism (Mudde and Rovira Kaltwasser, 2011; Stavrakakis, 2017). However, there are still pending questions in regards to the concrete articulation of populist discourse in its exclusionary and inclusionary variant, and especially about the ways whereby populist discourse is related to other discourses and sub-discourses. In this vein, an intriguing and relevant question, and the one concerning this paper, is how different representations and practical argumentations about Europe are integrated in and interact with populist discourses. Accordingly, this study conducts a comparative discourse analysis of two parties considered in the literature as inclusionary and exclusionary variants of populism: Podemos in Spain (Kioupkiolis, 2016; Sola and Rendueles, 2017) and Alternative für Deutschland (AfD) in Germany (Franzmann, 2016; Grimm, 2015).

The methodological approach of this research is highly influenced by the work of Normal Fairclough but explicitly relies also, in the fine methodological design, on corpus linguistics analyses, especially on the work of Paul Baker and associates (Baker, 2006; Baker and McEnery, 2015). Little CDA literature have devoted to the study of populism and populist discourse, with the notable exception of Ruth Wodak's works (2015, 2017), and there is still room to develop the fruitful synergy between CDA methodologies, especially corpus-assisted discourse analysis, and populism studies. The data collected in this study are electoral manifestos, press releases, and speeches of the parties corresponding to the period 2013 – 2017. Its systematic description, analysis and comparison are conducted with the software *Wordsmith 7.0* and a more nuanced discourse analysis is undertaken through *ATLAS.ti 8.0.* Results show concrete patterns of signification for each party derived from the connections, in the case of Podemos, between human right, populist and critical-Europeanist discourses and, in the case of AfD, from a combination of nationalist, populist and Eurosceptic discourses.

References

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

Baker, P. and McEnery, T. (eds) (2015) Corpora and Discourse: Integrating Discourse and Corpora. London: Palgrave.

Fairclough, N. (2013) Critical discourse analysis and critical policy studies, Critical Policy Studies, 7:2, 177-197

Fairclough, I. and Fairclough, N. (2012) Political Discourse Analysis. A method for advanced students. Routledge.

Simon T. Franzmann (2016): Calling the Ghost of Populism: The AfD's Strategic and Tactical Agendas until the EP Election 2014, German Politics.

Robert Grimm (2015) The rise of the German Eurosceptic party Alternative für Deutschland, between ordoliberal critique and popular anxiety. International Political Science Review, Vol. 36(3) 264–278.

Cas Mudde and Cristóbal Rovira Kaltwasser (2011) Voices of the peoples: populism in Europe and Latin America compared. Kellogg Institute.

Alexandros Kioupkiolis (2016) Podemos: the ambiguous promises of left-wing populism in contemporary Spain, Journal of Political Ideologies, 21:2, 99-120.

Jorge Sola and César Rendueles (2017): Podemos, the upheaval of Spanish politics and the challenge of populism, Journal of Contemporary European Studies.

Yannis Stavrakakis (2017) Discourse theory in populism research. Three challenges and a dilemma. Journal of Language and Politics. (Published online: 12 Jun 2017), 1–12.

Ruth Wodak (2015) The Politics of Fear: What Right-Wing Populist Discourses Mean. *SAGE*.

Ruth Wodak (2017) The "Establishment", the "Élites", and the "People" Who's who? *Journal of Language and Politics* 16:4 (2017), 551–565.

Construction grammar and academic discourse analysis

Nicholas Groom
University of Birmingham
n.w.groom@bham.ac.uk

Construction grammar was initially developed as a usage-based alternative to the nativist theoretical account of language propounded by Chomsky. This being the case, it is not surprising to observe that construction grammar remains to this day most strongly associated with cognitive linguistic theory and with psycholinguistic research into first and second language acquisition. In this paper, however, I aim to demonstrate that construction grammar offers equally exciting opportunities to corpus-based discourse researchers, particularly those whose work focuses on identifying and analysing the meanings and values associated with particular discourse communities.

The potential power of constructional approaches to corpus-based discourse analysis was first demonstrated by Wulff et al (2007), who identified statistically significant differences in the typical meanings (and thus the underlying cultural values) expressed by the '*into*-causative' construction in American and British English. The paper begins by considering why Wulff et al's call for further research along the lines set out in their paper has gone largely unheeded in the decade since their work was published. It is proposed that one of the main reasons for this neglect may be largely methodological; most current construction-based approaches are deductive in nature (i.e. the researcher decides in advance of the analysis which construction(s) to study), whereas discourse research is often exploratory in nature and thus more suited to inductive methods (i.e. where the aim is to discover, rather than make a priori assumptions about, which constructions are associated with a particular language variety or discourse community). The paper then proposes an adaptation of closed-class keywords analysis (Groom 2010) as a viable methodology for the inductive identification and analysis of discourse-specific

constructions in large computerized corpora.

The remainder of the paper will provide a practical illustration of this approach, showing how corpus-based construction grammar can yield new insights into the relationship between phraseology (defined as preferred ways of saying) and epistemology (defined as preferred ways of knowing) in the specialized discourses of academic disciplines. The main empirical focus of the paper will be on a newly identified construction, the **WAY IN WHICH** construction (as in *This may have affected the way in which religious ideas were disseminated*), and will draw on examples of this construction as it occurs in a large-scale corpus-based analysis of professional academic writing in the academic disciplinary discourses of history and literary criticism.

References

Groom, N. 2010. "Closed-class keywords and corpus-driven discourse analysis". M. Bondi and M. Scott (eds.) *Keyness in texts.* Amsterdam: John Benjamins, pp.59-78.

Wulff, S., Stefanowitsch, A. and Gries, S.T. 2007. "Brutal Brits and persuasive Americans: variety-specific meaning construction in the *into*-causative". In G. Radden, K. Köpcke, T. Berg and P. Siemund (eds.) *Aspects of meaning construction*. Amsterdam: John Benjamins. 265-81.

A diachronic move structure analysis of the patent specification genre, 1740 – 2011

Nicholas Groom
University of Birmingham
n.w.groom@bham.ac.uk

Jack Grieve
University of Birmingham
j.grieve@bham.ac.uk

Move structure analysis (MSA) is now a well-established and widely used methodology within the field of genre analysis (Tardy & Swales 2014). The aim of MSA is to identify conventional rhetorical structures within generic textual forms. The analyst begins by studying individual exemplar texts qualitatively, breaking each text down into a sequence of distinct and describable 'moves', understood as "stretch[es] of text serving a particular communicative (that is, semantic) function" (Upton & Cohen 2009: 588). Once this process has been iterated over a large number of exemplar texts within a given genre, the analyst can then aggregate their observations into generalizations about what seem to be the most common move sequence type(s) for that genre, and about which moves appear to be obligatory and which appear to be optional.

Whilst most MSA research has thus far been synchronic in orientation, there is now growing interest in the question of how and why particular moves and move structure configurations appear, develop, change and/or disappear over time. Our paper contributes to this emerging research frontier by carrying out the first ever diachronic analysis of a historically significant, and yet currently under-researched, legal/technical genre: the patent specification. Our analysis focuses on a corpus of 272 British patent specifications ranging from the eighteenth century to the present day. The methodology of the study involves reducing exemplar texts to simple code strings representing individual move sequence configurations, which we then process using string edit distance (Navarro 2001), a technique which is commonly used in natural language processing and genetics but which has to our knowledge never been used in genre analysis before.

The main findings of our study are that while patent move sequence types usually change very gradually, many individual moves within these sequence types change very abruptly; they do not usually appear or disappear gradually over a number of years, as might be expected. The explanation for this somewhat counter-intuitive pair of observations is that individual moves fall in and out of usage at different points in time, thereby leading to a more gradual overall change profile. We conclude by considering the broader significance of our study for current theoretical debates about the nature of genre change. We argue that our results lend support to the 'evolutionary' model of genre change proposed by Gross et al (2002), but also that, somewhat paradoxically, this gradual evolutionary process is driven by very sudden changes at the individual move level.

References

Gross, A. G., Harmon, J. E., and Reidy, M. 2002. *Communicating science. The scientific article from the 17th century to the present*. Oxford: Oxford University Press.

Navarro, G. 2001. "A guided tour to approximate string matching". *ACM computing surveys (CSUR)*, 33(1). 31-88.

Tardy, C.M. and Swales, J.M. 2014. "Genre analysis". In K.P. Schneider & A. Barron (eds.) *Pragmatics of discourse*. Berlin: Walter de Gruyter, pp.165-187.

Upton, T.A. and Cohen, M.J. 2009. "An approach to corpus-based discourse analysis: The move analysis as example". *Discourse Studies* 11(5): 585–605.

Mediating *Truth*, *Fact* and *Reality* through Metadiscourse: A Corpus-based CDA Analysis of Government-affiliated Interpreters' Agency at China's Political Press Conferences

Chonglong Gu
The University of Manchester
chonglong.gu@manchester.ac.uk

The interpreter-mediated and televised Premier-Meets-the-Press conferences are an institutionalised annual event held within the broader context of reform and opening-up in mainland China (Gu 2018). A vital site of ideology, this discursive event offers the Chinese premier the opportunity to articulate China's policies and official positions and construct a favourable image through answering journalists' questions on a wide range of issues (e.g. economic and political restructuring, anti-corruption efforts, global financial crisis, Taiwan, Tibet, and the Korean Peninsula). Unsurprisingly, a key component of the press conferences involves the discursive articulation of China's official version of *truth*, *fact* and *reality*, given the essentially political and ideological nature of the event. With this in mind, this corpus-based CDA study concerns an examination of the government-affiliated interpreters' agency and, more specifically, their metadiscursive mediation and (re)construction of China's political discourse relating to truth, fact and reality.

While truth value is very much embedded and conveyed in the main propositional content, given the limited space, attention is focused specifically on the metadiscursive items that are explicitly related to fact, truth and reality semantically. Metadiscourse refers to 'the linguistic material which does not add propositional information but which signals the presence of an author' (Vande Kopple 1985) or 'the means by which propositional content is made coherent, intelligible and persuasive to a particular audience' (Hyland 2005: 39). Although metadiscourse

might not automatically add to the actual propositional content, it represents the text producer's 'intrusion into the discourse, either explicitly or non-explicitly' (Crismore 1983: 2) and it can indeed be 'ideology-bound' and relate 'significantly to the distribution of power in a given context' (Mazid 2014: 76). So far, metadiscourse has been mostly explored in various academic writing genres (e.g. Jiang and Hyland 2015). However, it has received little, if any, scholarly attention in (interpreted) political discursive articulation – a dynamic area featuring a strong level of ideological legitimation and justification – from the perspective of discourse and ideology.

As such, drawing on the CE-PolitDisCorp corpus established by the author, which contains 20 years of press conference data (1998-2017), eligible metadiscursive items in the interpreted English subcorpus (e.g. *in fact*, *as a matter of fact*, *truly*, *really*, *the fact that*, *actually*, *indeed*) and eligible items in the corresponding Chinese subcorpus are established and counted. Statistically, it is found that the interpreters tend to (over)produce these items in English. Therefore, given the semantic meanings of these items relating to *fact*, *truth* and *reality*, a stronger and more emphatic sense of factualness, truthfulness and authority are (re)created in the interpreted English discourse overall.

For more refined and detailed CDA analysis, further attention is focused on the metadiscursive noun expression 'the fact that' as an illustrative example. Due to the linguistic property to front-load attitudinal meanings and express stance, this specific structure is proven useful in knowledge construction, for instance, in academic writing (e.g. Jiang and Hyland 2015). This metadiscursive noun structure, however, remains largely underexplored in interpreter-mediated political encounters. Critical comparative analyses between the Source Text (ST) and Target Text (TT) reveal that the interpreters tend to proliferate the use of the structure "*the fact that*" in the English discourse. Discursively, this not only adds an additional layer of factualness and authority to the Chinese premier's discourse but also leads to the further (re)construction of positive *self* representation (e.g. China's tremendous achievements are due to "the fact that we [the Chinese government] have always upheld unity of the Party and safeguarded social and political stability") and negative *other* representation in interpreting (e.g. "the main problems in China-Japan relations lie in the fact that some leaders in Japan keep on visiting the Yasukuni shrine"). Inherently interdisciplinary in nature, this study, focusing on interpreters' agency in the metadiscursive reconfiguration of reality, promises to further contribute to scholarship in interpreting studies, CDA, corpus linguistics and Chinese studies alike.

References

Crismore, Avon (1983) *Metadiscourse: What It is and How It is Used in School and Non-School Social Science Texts*, Center for the Study of Reading (Vol. Technical Report No. 273) Urbana Champaign: University of Illinois.

Gu, Chonglong (2018) "*Towards a Re-Definition of Government Interpreters' Agency Against a Backdrop of Sociopolitical and Cultural Evolution: A Case of Premier's Press Conferences in China*". In Olaf Seel (eds.) Redefining translation and interpretation in cultural evolution. Hershey, PA: IGI Global.

Hyland, Ken (2005) *Metadiscourse: Exploring interaction in writing*, London and New York: Continuum.

Jiang, Feng and Ken Hyland (2015) ''The Fact That: Stance nouns in disciplinary writing", *Discourse Studies* 17(5): 529-550.

Mazid, Bahaa-eddin M. (2014) *CDA and PDA Made Simple: Language, ideology and power in politics and media*, Newcastle upon Tyne: Cambridge Scholars Publishing.

Vande Kopple, William J. (1985) 'Some Exploratory Discourse on Metadiscourse', *College Composition and Communication* 36: 82-93.

In, out and in-between: a corpus-assisted exploration of the identities of dwellers, foreigners and travellers in EModE

Caterina Guardamagna
University of Liverpool
C.Guardamagna@liverpool.ac.uk

Ursula Kania
University of Liverpool
Kania@liverpool.ac.uk

Victorina González Díaz
University of Liverpool
vgdiaz@liverpool.ac.uk

This study, developed in dialogue with the *Keyword* component of the ERC-funded University of Liverpool *Travel, Transculturality and Identity in England* 1550-1700 (TIDE) research project, focusses on a subset of words belonging to the *Government and Law* pathway, namely *citizen, denizen, traveller, vagrant, settler, foreigner, alien* and *stranger*. Due to the technical impossibility of identifying keywords on statistical grounds (*keyness*), these terms have been identified in consultation with the TIDE team of cultural historians who, taking an approach akin to Williams (1976), defined them as particularly relevant in Early Modern debates.

This study focuses on the 1550-1700 period in the light of both earlier and later developments investigated through a combination of data from of the OED and corpus exploration (EEBO, Helsinki and the PENN corpus collection). A corpus assisted discourse analytical approach (Stubbs 2001, Baker 2006) -- relying mainly on frequency of usage, collocation analysis, and semantic prosody – resulted in the compilation of a "typical" profile for each of the term considered. These "profiles", corresponding to the discursive representations of these groups, are then systematically compared and contrasted. For instance, whereas *vag(a)rant* does not undergo any major semantic shift in EModE, keeping the negative connotation of a disreputable, dishonest, idle individual paired up or identified with beggars, rovers, vagabonds and rogues, mixed attitudes towards travellers develop in the period considered to be then sustained over time. Specifically, while unmodified forms appear in context suggesting a negative evaluation (e.g. *beware the traveller*) because of the tall tales and lies travellers bring from afar, premodification suggests that travellers are seen as adventurers/explorers, even celebrities/heroes acting as sources of knowledge and progress (e.g. *ingenious/judicious traveller*). The difference in the representation of these two groups is evident also in the fact that they share only one of their top-50 collocates in EEBO. Further semantic developments of *traveller* include the professionalization of the term (OED, end of the 18th c), which is also found in related terms like e.g. *ambassador*.

Discursively, three main categories are identified based on the aforementioned profiles: dwellers (settlers, citizens, denizens), foreigners (foreigners, strangers, aliens), movers (travellers, vagrants). These are not fixed but interact with each other in a complex interplay of in-group and out-group memberships (a clear case in point is that of *denizen* and also *settler*) while also interacting with other figures with whom they share some features (e.g. soldiers, pilgrims, merchants, mariners).

This paper contributes to a growing field of diachronic corpus assisted discourse studies (CADS), extending the methodology used for modern approaches (e.g. the ESRC RASIM research project by Baker et al. carried out at Lancaster University and Partington 2010) to historical studies as demonstrated by Culpeper and Findlay (2014).

References

Baker, P. 2006. *Using Corpora in Discourse Analysis*. London: Continuum.
Baker, P., McEnery T., Wodak, R., Gabrielatos, C., KhosraviNik, M., and Krzyzanowski, M. 2007. *Discourses of refugees and asylum seekers in the UK press, 1996-2006*, Lancaster University.

Culpeper, J. and Findlay, A. 2014. "Contemporary Understandings of Welsh, Scottish and Irish identities: Celtic Characters in Shakespeare's Henry V". paper delivered at *The Northern Renaissance Seminar in Association with CREME*, 22nd of February 2014, Lancaster University.

Partington, A. 2010 *Modern Diachronic Corpus Assisted Discourse Studies*, Edinburgh: Edinburgh University Press.

Stubbs, M. 2001. *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Investigating Opposition: A Corpus and Critical Stylistic Study of Religion in Broadsheet Newspapers

Sawsan Hazim Hassan
University of Huddersfield
Sawsan.Hassan@hud.ac.uk

Although there is a great interest in studying religion from a different range of disciplines, including media, sociology, theology and anthropology, investigating the representation of religion from a linguistic perspective remains limited. In this presentation, I will talk about three widespread religions in the UK – Judaism, Christianity and Islam, focussing particularly on the groups, namely: Christians, Muslims and Jews to show how oppositions are constructed in British broadsheet newspapers. The purpose of this study is to show how conventional oppositional pairs can be exploited in the creation of new, non-canonical opposites of each participant for ideological gain. The aim is to explore whether there is a difference, or even a similarity, in the representation of the three participants within the period of time under study, and whether the three religions are portrayed in a balanced way.

The corpus built for this study was collected from all Broadsheet newspapers published in 2010. It is a collection of 12,580 articles, which contain the words *Christian*(s), *Christianity*, *Muslim*(s), *Islam*, *Islamic*, *Jew*(s), *Judaism* and *Jewish*. However, the instances of contrast used for the purposes of the analysis were selected through the examples captured for naming strategies and transitivity analysis for relational processes of this research. This made my search more manageable as it provides good evidence that reveals patterns for ideological construction. The study was conducted using the corpus software *Sketch Engine* (Kilgarriff et al. 2004) and the model of contrasting offered by Jeffries (2010a) and Davies (2008; 2012) to identify any differences or similarities among the three religions. *Sketch Engine*, for example, provides significant collocates based on the grammatical relations among words within the corpus. According to Davies, understanding new, non-conventional opposites often relies on understanding the superordinate canonical opposites (2008). This can be helpful to note any higher level oppositional concepts that were present in the textually constructed examples. A detailed qualitative analysis of textual construction of opposition was conducted to show how opposites are used by text producers to represent a group of individuals positively or negatively in order to maintain a particular ideological point of view. The analysis is carried out across several levels, including structural and semantic features. The structural or syntactic triggers of opposition include negation, transitional opposition, comparatives, replacive opposition, concessive, contrastive and parallel structure. References to semantic features depend on the meanings of some of the lexical choices in the context (Jeffries, 2010b).

In order to examine the representations of the three religions in the broadsheet newspapers, the corpus was subjected to a comparative analysis by examining the oppositional pairs that

contribute to portray each group in a different way. The quantitative and qualitative analyses helped to explain the findings which show, for example, different and similar superordinate oppositions in the representation of Christians, Muslims and Jews. I found that one of the significant structuring devices used in the texts about the three participants was the construction of unconventional opposites of the kind I am inspecting here. What was most remarkable about this part of the study was the way in which the majority of the textually-constructed opposites were related either implicitly or explicitly to the three conventional superordinate concepts: *positive/negative* or *belief/disbelief* pairs. Additionally, the oppositional patterns show a set of syntactic triggers, which are associated with a certain group rather than another. Parallelism and contrastive construction, for instance, are associated with the representation of Muslims, while it is negation and contrastive opposition that are salient in the construction of Jews and Christians. This helps to show a focus around the subject of bias, in particular, positive or negative biases.

References

Baker, Paul (2006) *Using corpora in discourse analysis*. London: Continuum International Publishing Group Ltd.

Baker, P., Gabrielatos, C., & McEnery, T. (2013) *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press.* Cambridge: Cambridge University Press.

Davies, M. (2008) *Oppositions in News Discourse: the Ideological Construction of Us and Them in the British Press*. Unpublished PhD thesis, University of Huddersfield.

Davies, M. (2012) *Oppositions and Ideology in News Discourse*. London: Bloomsbury.

Jeffries, L. (2010a) *Critical Stylistics*. Basingstoke: Palgrave Macmillan.

Jeffries, L. (2010b) *Opposition in Discourse*. London: Continuum.

Stubbs, M. (2001) *word and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.

Construction of male and female identities in the 'new misogyny': a corpus-based discourse comparison of a misogynistic murderer's manifesto and misogynistic online forums

Abi Hawtin
Lancaster University
a.hawtin@lancaster.ac.uk

On 23rd May 2014, 22 year old Elliot Rodger killed 6 people and injured 13 others in California. He left behind a series of YouTube videos in which he discussed his hatred of women, and a 'manifesto' which outlined his life up to that point, his views on women, and his plan to take revenge. In the first part of this study I use corpus methods (collocation and semantic collocation analysis) to analyse the ways in which Rodger constructs the identities of men and women in his manifesto. Rodger is perhaps unusual in that he wrote at sufficient length prior to his murders that the resulting document ('My Twisted World: The Story of Elliot Rodger') is large enough for corpus analysis by itself. Although commonly referred to as a 'manifesto', it contains an autobiographical account with a particular focus on his relationship with women and his plans to punish them.

The first type of analysis I performed was a collocation analysis, using AntConc (Anthony, 2017). I subsequently used Wmatrix (Rayson, 2009) to conduct a similar collocation analysis, but using USAS tags and exploiting Wmatrix's ability to search for semantic tag collocates.

Looking at the collocates of the female and male search terms helps to reveal several discourses which contribute to the ways that Rodger constructs the identities of men and women in his manifesto. Primarily, he constructs both men and women (i) as homogenous groups which he is outside of, and (ii) in terms of their appearance. Rodger furthermore constructs women as objects of his hatred, targets for his revenge, as goals which he wants to achieve, but also as having power over him and men in general. By contrast, he constructs men as a group as being able to have experiences, usually sexual, which he cannot.

All these tags contribute to the expression of a discourse of women having the power to 'choose' men. However, it becomes clear when looking at the concordance lines that Rodger feels that women are using this power wrongly by choosing the 'wrong' men. This discourse of 'unjust power' is by far the most prominent element of Rodger's representations of men and women, with almost all of the collocations (with both words and semantic tags) contributing in some way towards his construction of women as having a great amount of power, but using this power wrongly.

The analysis above shows that Elliot Rodger consistently expresses and constructs an ideology that in contemporary media commentary is commonly called the 'new misogyny'. This is a worldview in which men view women as privileged and powerful, and themselves as oppressed. This kind of sexism stands in contrast to a more traditional patriarchal ideology.

In order to investigate just how similar Rodger's views were to others who identify as believing in this worldview, I carried out a similar corpus-based discourse analysis of the ways in which men and women are constructed on internet forums which are popular with men's rights activists, in order to show whether the ways in which Rodger viewed men and women are characteristic of this new and more dangerous type of misogyny which is becoming increasingly prevalent online. I find that many of the same discourses surrounding men and women are present in these online forums. This highlights just how dangerous these forums can be, and shows that the 'echo-chamber' nature of them could act as encouragement for people to commit such terrible acts as those committed by Rodger.

References

Anthony, L. (2017). AntConc (Version 3.5.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software.

Rayson, P. 2009. Wmatrix: a web-based corpus processing environment, Computing Department, Lancaster University. Available at: http://ucrel.lancs.ac.uk/wmatrix/

Gap between Rhetoric and Reality: China's Self-branding of its Rise and the Perceptions in British and American Press

Zeping Huang
Hong Kong Baptist University
huangzeping@gmail.com

Yunya Song
Hong Kong Baptist University
songyunya@gmail.com

China has experienced sustained economic growth since its reform and opening-up in the late 1970s. By the middle of 2010, its gross domestic product was reported to have surpassed Japan's GDP, making China the second largest economy in the world (Gustin, 2010). In recent decades,

the topic of a rising China has attracted worldwide attention as well as criticism. Its increasing use of soft power to promote its foreign policy interests has been a focus of Western media coverage and policy discourse (Barr, 2011; Ding, 2010). Notably, China's perception of itself appears strikingly at odds with that of others (Brown & Loh, 2011). On the one hand, China's national image benefits from its strong commitment to cultural diplomacy, effective employment of information and communication technologies in international broadcasting, and close connections with the Chinese diaspora around the world (Ding, 2011). Beijing's self-congratulatory slogans—ranging from former president Hu Jintao's "harmonious society" and "scientific concept of development" to President Xi Jinping's "Chinese dream" and "Chinese renaissance"—have projected a coherent view of China as a peaceful country with a vision of building a prosperous, strong, harmonious and modernised socialist society (Buzan, 2014; Callahan, 2015). On the other hand, its national image management is constrained by such hurdles as weak political credibility, reviving popular nationalism, and various missteps in its foreign policies (Ding, 2011). Against this backdrop, the study draws upon a Discourse Historical Approach (Reisigl & Wodak, 2009) to corpus-assisted studies (Baker et. al, 2008) in an attempt to investigate discrepancies in the discursive strategies employed by the UK, China and U.S press coverage of China's rise. Two primary questions are formulated as follows:

1. In what way and to what extent do Chinese newspapers differ from British and American newspapers in reporting China's rise?
2. Do similarities and differences exist between British and American newspapers in reporting China's rise? If so, what are they?

To address the above questions, this paper delves into the following three sub-questions.

a). How is the topic of *China's rise* conceptualized in the news discourse of these three countries?

b). How is *China's rise* characterized in the three countries' news discourse?

c). From what perspectives are the representations of China's rise expressed in the news discourse of the three countries?

The data used in the study is a 2.5 million-word corpus composed of news texts containing three search phrases (*China's rise, rise of China, emergence of China*) in six newspapers between 1 January, 2009 and 31 December, 2017. The six newspapers are *China Daily* and *Global Times* from China (hereafter referred as China corpus), *The Times* and *The Guardian* from the U.K. (UK corpus), *The Washington Post* (WP) and *New York Times* (NYT) from the U.S. (US corpus). With the help of the corpus tool WordSmith 5.0 (Scott, 2012), this comparative study focuses on keyword and collocation analysis. First, keywords are generated by comparing the China, UK, and US corpora against one another. The top 100 keywords are compared according to semantic categories. Based on the frequency and MI (mutual information) scores, the top 100 collocates of the three search phrases (from L1 to L5 and R1 to R5) are also identified from the three corpora respectively. Lexical noun, adjective, and verb collocates are selected for close examination of their contextual use through concordance analysis. Following the concordance analysis, the collocates are then classified into four categories of discursive strategies: normalization, predication, argumentation and perspectivation. Both quantitative and qualitative results will be presented and discussed.

References

Baker, P., Gabrielatos, C., Khozravinik, M., McEnery, T., & Wodak, R. 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses on refugees and asylum seekers in the UK press. Discourse & Society 19 (3): 273-305.

Barr, M.T. 2011. *Who's afraid of China?: the challenge of Chinese soft power*. New York; London: Zed Books.

Brown, Kerry and Loh Su Hsing (2011), *Trying to read the new 'assertive' China right*, Asia Programme Article: ASP PP 2011/02, London: Chatham House.

Buzan, B. 2014. The logic and contradictions of 'peaceful rise/development' as China's grand strategy. *The Chinese Journal of International Politics* 7(4): 381-420.

Callahan, W.A., 2015. Identity and security in China: The negative soft power of the China dream. *Politics* 35(3-4): 216-229.

Ding, S. 2010. Analyzing rising power from the perspective of soft power: a new look at China's rise to the status quo power. *Journal of Contemporary China* 19(64): 255-272.

Ding, S. 2011. Branding a Rising China: An Analysis of Beijing's National Image Management in the Age of China's Rise. Journal of Asian and African Studies.

Gustin, S. 2010. Now the No. 2 economy, what's ahead for China? Retrieved in Aug. 2010, Daily Finance http://www.dailyfinance.com/story/China-number-two-whats-head/19595235/

Reisigl, M. & Wodak, R. 2009. "The discourse-historical approach (DHA). In R. Wodak, M. Meyer (eds.) *Methods for critical discourse analysis*. London: Sage.

Scott, S. 2012. WordSmith Tools (Version 5.0) [Computer Software].

Collocation analysis to identify the 'unusual' in hybrid texts

Reka R. Jablonkai
University of Bath
r.r.jablonkai@bath.ac.uk

The texts produced in multicultural and multilingual contexts, for example, institutions of the European Union and the United Nations, are often referred to as hybrid texts. The definition proposed by Trosborg (1997) says that hybrid texts: "are arrived at as an outcome of negotiations between cultures and the norms and conventions involved as well as through translation" (p. 146) and exhibit "features that somehow seem 'out of place'/'strange'/'unusual' for the receiving culture" (Schäffner & Adab, 2001, p. 175). Written documents produced within the EU institutional context can be considered hybrid texts and this hybridity might be reflected in the collocational patterns and semantic preferences of specific lexical items in this specific discourse (Gledhill, 2000; Mautner, 2016; Nelson, 2006). The present paper reports on the collocation analysis of English EU discourse that aimed to identify discourse-specific patterns of collocations of 16 frequent lexical items in written English EU documents. The study investigated the English EU Discourse corpus, which is a one-million-word corpus of English EU documents designed and compiled based on the results of a needs analysis among EU professionals. Collocational patterns were identified with the help of Sketchengine (Kilgarriff & Tugwell, 2001). The study was guided by the following research questions: RQ1 What collocational patterns emerge in written English EU discourse? RQ2 How do patterns in the EEUD corpus compare to collocational patterns in the BNC Written?

The analysis revealed that there are differences in the collocates and grammatical relations the selected lexical items frequently form in the EEUD Corpus and in the BNC Written. The different collocates or frequent untypical grammatical behaviour of lexical items might be perceived as 'strange' or 'unusual' features of EU texts. Therefore, these features can be considered as elements contributing to the hybridity of English EU documents. The number of collocates, and the variety of semantic preferences the selected lexical items appear to have, is greater in the BNC Written, which suggests a certain degree of fixedness in the lexical aspects of collocation in the EEUD Corpus.

References

Gledhill, C. (2000). "The discourse function of collocation in research article introductions". *English for Specific Purposes, 19*(2), 115-135.

Mautner, G. (2016). "Checks and balances of critical discourse studies". In R. Wodak and M. Meyer (eds.) *Methods of critical discourse studies.* London: SAGE.

Nelson, M. (2006). "Semantic association in Business English: A corpus-based analysis". *English for Specific Purposes, 25*(2), 217-234.

Kilgarriff, A., & Tugwell, D. (2002). "Sketching words". In M. Corréard (Ed.), Lexicography and natural language processing: A festschrift in honour of B. T. S. Atkins (pp. 125-137). EURALEX. Retrieved August 25, 2009, from http://www.kilgarriff.co.uk/publications.htm

Schäffner, C., & Adab, B. (2001). "The idea of the hybrid text in translation: Contact as conflict". *Across Languages and Culture, 2*(2), 167-180.

Trosborg, A. (1997). "Translating hybrid political texts". In A. Trosborg, (Ed.), *Text typology and translation* (pp. 145-158). Amsterdam: John Benjamin Publishing Company.

"And then she called the GP…": a corpus-assisted discourse study of gender and the language of physical pain in chronic and terminal illness

Sylvia Jaworska
University of Reading
s.jaworska@reading.ac.uk

Drawing on the notion of gender as a socially constructed category performed inter alia through repeated and regulated language use (cf. Butler 1990; Baker 2014), this study examines the ways in which women and man use language to do person-in-pain in real-life interactions about chronic and terminal illness. This study is based on a secondary analysis of a large corpus of illness narratives collected by the Health Experiences Research Group at the University of Oxford and published by the DIPEx Charity. Eighteen conditions were identified in which men and women talked about physical pain. Two separate gender-based corpora were created MCor for men containing 107,283 words and WCor for women containing 135,433 words, and interrogated using Sketch Engine. The lemma 'pain' and other descriptors of pain were analysed for frequency of use. 'Pain' emerged as the most frequent lexical choice to refer to pain experience and hence, it was further examined by studying its collocations to reveal the kinds of meanings and associations that women and men attached to pain.

Our study shows significant quantitative and qualitative differences in the ways in which women and men report pain pointing to the existence of distinctive feminine and masculine lexical repertoires of pain talk. These repertoires reflect differentiated approaches to pain, some of which appear to be underpinned by gender stereotypes. Women refer to pain more frequently and have a wider lexical repertoire for pain reporting. They also use more references to psychological actions around pain suggesting that the female experience of physical pain is more likely to be perceived as psychosomatic. In contrast, men tend to use fewer descriptors, most of which are highly emotive words suggesting that they report pain when it becomes unbearable. There is also a conspicuous absence of references to psychological processes in the male narratives and the focus is on pain management through pain killers. Understanding these differences can help health professionals respond effectively to people's talk about pain and develop more holistic and gender-aware practices in pain consultation leading potentially to the reduction of gender biases and inequalities in healthcare. This paper contributes to the growing

body of corpus-based research in health communication (e.g. Harvey 2012; Semino et al. 2015; Potts and Semino 2017; Brookes and Baker 2017; Jaworska and Kinloch 2018) and highlights the benefits of corpus linguistic insights to a better understanding of lay experience of health and illness.

References

Brookes, G. and Baker, P. 2017. What does patient feedback reveal about the NHS? A mixed methods study of comments posted to the NHS Choices online service. BMJ Open 201.

Butler, J. 1990. Gender trouble: Feminism and the subversion of identity. New York, Routledge.

Harvey, K. 2012. Disclosures of depression: using corpus linguistics methods to interrogate young people's online health concerns. International Journal of Corpus Linguistics. 17(3), 349-379.

Jaworska, S. and Kinloch, K. 2018. Using multiple data sets. In: Taylor, C. and Marchi, A. (eds.) Corpus Approaches to Discourse: A Critical Review. Routledge, London: 110-129.

Doing CADS with multiple data sets: the case of postnatal depression (PND)

Sylvia Jaworska,
University of Reading,
s.jaworska@reading.ac.uk

Karen Kinloch
Edge Hill University
kinlochk@edgehill.ac.uk

Taking up the claim made by Partington et al. (2013: 12) that 'we are not deontologically justified in making statements about a relevance of a phenomenon observed in one discourse type unless […] we compare how the phenomenon behaves elsewhere', our paper explores the methodological potential of using multiple data sets in corpus-assisted discourse studies (CADS).

Previous work in CADS has employed researcher (Marchi and Taylor, 2009; Baker and Egbert, 2016) and methodological (Baker and Erez, 2015) triangulation as ways of reducing partiality and "cherry picking" but as yet there have been few instances of data triangulation in this field.

We adopt the linguistic methodology of a comparative corpus-assisted discourse study and centre our analysis on a comparison of keywords and their collocational profiles (Partington et al. 2013). Specifically, we show how a comparative corpus-based analysis of textual data on the same topic but produced in different professional and institutional domains can highlight shared as well as divergent meanings used to construct a discourse type. We also look at how these are recontextualised, appropriated or excluded across the selected domains.

We exemplify this approach by exploring discursive constructions of postnatal depression (PND) across medical, media and lay contexts. Our data sets include four textual domains: 1) lay narratives sourced from Mumsnet, 2) documents about PND disseminated by clinicians for clinicians, 3) information by clinicians for lay people and 4) articles about PND from British newspapers. The focus of this paper is the keyword PND/postnatal depression, which is examined using collocates, concordances and WordSketches (Kilgarriff et al. 2004) from each of the textual domains.

This comparative approach allows us to identify differences, similarities and absences in the discursive resources which are adopted to "talk about" PND across the different contexts. Our results show how metanarratives, especially the dominant biomedical model of health and

illness and contemporary mothering ideologies (Pedersen 2016) are strategically appropriated and contested by women with PND to reduce felt stigma, assume agency and repair 'spoilt identity' (Goffman 1963).

At the methodological level, our study highlights the relevance of using a comparative discourse approach to foster our understanding of the role that social and medical discursive resources play in constituting the lay experience of health and illness. This case study will also demonstrate the ways in which the discursive construction of a particular condition varies across data sets from four different text producers and how such multidimensional and comparative research can reduce partiality and enrich analysis of a given topic highlighting also some potential difficulties which arise from this approach.

References

Baker, P. and Egbert, J. (eds) (2016). *Triangulating Methodological Approaches in Corpus-Linguistic Research*. London: Routledge.

Baker, P. and Erez, L. (2015). Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity. In: *Discourse & Communication* 9(2): 221–236.

Goffman, E. (1963). *Stigma. Notes on the Management of Spoiled Identity*. Englewood Cliffs: Prentice-Hall.

Kilgarriff, A., Rychlý, P., Smrz, P. and Tugwell, D. 2004. The Sketch Engine. *Proc EURALEX* 200. Lorient, France, 105–116.

Marchi, A. and Taylor, C. (2009). If on a winter's night two researchers … A challenge to assumptions of soundness of interpretation. In: *CADAAD Journal 3(1):* 1–20.

Partington, A., Duguid, A. and Taylor, C. (2013). *Patterns and Meanings in Discourse. Theory and Practice in Corpus-Assisted Discourse Studies* (CADS). Amsterdam: Benjamins.

Pedersen, S. (2016). The good, the bad and the 'good enough' mother on the UK parenting forum Mumsnet. In: *Women's Studies International Forum* 59: 32-38.

Negotiation of hybrid identities in a BBC sitcom: Muslim, British and Pakistani.

S. Bilal Kadiri
King Khalid University
skadiri@kku.edu.sa

There are over 1.17 million Pakistanis living in the UK (Office for National Statistics 2011) and over 90% of them identify themselves as being Muslim. However, the identities of British Pakistani Muslims should not be considered as a monolithic entity, but rather as a hybrid of intersecting components that include race, religion, nation, age and gender. For example, the identity of a 19 year old British born Pakistani Muslim woman would differ to that of her 75 year old grandmother, who may have spent most of her life in Pakistan.

This paper addresses the different ways members of this group negotiate their hybrid identity by drawing upon example data from the CK corpus (40,000 words). This corpus contains transcripts of thirteen episodes of the BBC sitcom *Citizen Khan*, which is centred on the lives of a British Pakistani Muslim family. The British press have frequently associated Muslims with discourses linked to terror and extremism (Baker et al. 2013, Moore et al. 2008), resulting in the Muslim community being looked at with a hint of suspicion from some quarters of the wider British Public, especially in relation to how well they assimilate into British society and how strongly they associate with their 'British identity'. Aside from discourses related to identity,

this paper additionally contributes towards the wider body of research examining 'media representations of Islam' (e.g. Baker et al. 2013, Richardson 2004). *Citizen Khan* was broadcast on a primetime slot on BBC One to a majority non-Muslim audience, affording the opportunity for many in rural parts of Britain who had limited interactions with Muslims to gain an insight into the lives of British Muslims.

My presentation will demonstrate how corpus software was used to identify salient aspects of identity within the sitcom and how upon further analysis it was determined that the characters negotiate specific aspects of their hybrid identity (i.e. religious identity) in a unique manner to one another. I will also point out how the scriptwriters have used such instances to generate humour and how my findings in this regard, are further supported through the personal perspectives of one of the writers. Additionally, it will be highlighted that a character may be afforded power over other characters due to their (mis)understanding of a particular facet of their hybrid identity. Thus, building upon Crenshaw's (1989) *theory of intersectionality*, which notes that power relations and inequalities can be compounded due to intersections between different types of disempowered identities (e.g. a black woman and white woman will not have the same experience of femininity). Finally, the paper will detail some of the issues that may arise when there are 'blurred lines' between the intersecting components of a hybrid identity (i.e. being able to distinguish between Pakistani identity and Muslim identity).

References

Baker, P., Gabrielatos, C. and McEnery, T. 2013. *Discourse Analysis and Media Attitudes: The representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Crenshaw, K. 1989. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *University of Chicago Legal Forum: Vol. 1989. Issue 1*. Article 8.

Moore, K., Mason, P., and Lewis, J. M.W. 2008. "Images of Islam in the UK: the representation of British Muslims in the national print news media 2000-2008". [Working Paper]. Cardiff: Cardiff University. Retrieved on 5 January 2018 from: http://orca.cf.ac.uk/53005/

Office for National Statistics. 2011. "2011 Census: Ethnic group, local authorities in the United Kingdom". Retrieved on 3 January 2018 from: http://www.ons.gov.uk/ons/rel/census/2011-census/key-statistics-and-quick-statistics-for-local-authorities-in-the-united-kingdom---part-1/rft-ks201uk.xls

Richardson, J. E. 2004. *(Mis) representing Islam: The racism and rhetoric of British Broadsheet Newspapers*. Amsterdam: John Benjamins.

Marriage for all ('Ehe fuer alle')?! A corpus-assisted discourse analysis of the equal marriage debate in Germany

Ursula Kania
University of Liverpool
ursula.kania@liverpool.ac.uk

In August 2001, Germany introduced 'civil unions' ('eingetragene Lebenspartnerschaften') for same-sex couples. Up until June 2017, the law had been revised several times, in part driven by rulings made by the Federal Constitutional Court. However, 'civil union' and 'marriage' were still two separate legal institutions, with certain rights (like joint adoption and tax benefits) only granted to married couples. After numerous initiatives proposing to open marriage to same-sex couples, the relevant bill finally passed on June 30th, 2017– the snap vote was triggered by

chancellor Merkel's unexpected move towards a 'vote of conscience' on this issue. The law came into force on October 1st, 2017.

The current study is situated within corpus-assisted discourse analysis (e.g., Baker & McEnery 2015) and aims at identifying key topics and attitudes in the marriage equality debate in Germany by providing quantitative and qualitative analyses of two corpora which were created using *Nexis*. They include texts from three German newspapers (*TAZ*, *Welt/Welt am Sonntag, Frankfurter Rundschau*) and two magazines (*Der Spiegel, Bunte*) discussing same-sex marriage/civil unions.

Corpus I covers the period from July 2000 (when the bill about 'civil unions' was first discussed in the German Bundestag) to August 2001 (when the first ceremonies took place); Corpus II covers the period from May 2016 to June 2017 (including coverage of the general election campaign, with marriage equality as one of the hot topics, culminating in the snap vote on marriage equality).

Analyses using *AntConc* (Anthony 2014) and *ProtAnt* (Anthony & Baker 2015) suggest that there are strong parallels to marriage equality debates in other countries (e.g., Vigo & Milani 2017 on Italy, Turner et al. 2017 on the UK), e.g., the notion of same-sex marriage posing a threat to society by undermining the institution of marriage (i.e., the so-called 'slippery slope' argument) is prevalent in the German data as well.

However, there are also language-/culture-specific features. For example, German seems to possess a particularly rich inventory of terms referring to legally recognized same-sex unions (such as 'Homo-Ehe', 'Schwulen-Ehe', 'gleichgeschlechtliche Ehe', 'Ehe fuer alle'). There are also clear longitudinal trends, as evidenced for example by a stronger focus on definitions of 'marriage' and 'family' as well as children's rights in the more recent data. Through an exploration of (changes in) frequency, collocates, and semantic prosody, it is shown how lexical choices contribute to framing the arguments for/against marriage equality in particular ways and how these stances are connected to wider German cultural norms.

While the corpus-based study has been completed, the collection/analyses of other types of data is still ongoing. In my presentation, I will address some of the issues that arise when trying to incorprate other sources of information (such as the images used in the newspapers and campaign ads used in the run up to the latest general election).

Overall, this research project aims at contributing to the growing body of CADS research in two ways: 1. by providing an analysis of a topic area that – to the author's knowledge - has not been explored so far (i.e., the marriage-equality debate in Germany); 2. by explicitly addressing methodological issues (such as the incorporation of other types of data).

References

Anthony, L. 2014. *AntConc* (3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Anthony, L. and Baker, P. 2015. *ProtAnt* (Version 1.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baker, P. and McEnery, T. (eds.) 2015. *Corpora and Discourse: Integrating Discourse and Corpora*. London: Palgrave.

Turner, G., Mills, Sara, van der Bom, I., Coffey-Glover, L., Paterson, L. L., Jones, L. 2018, Opposition as victimhood in newspaper debates about same-sex marriage, *Discourse and Society* 29(2).

Vigo, F. & Milani, T. M. (2017). *Italy and its civil unions: a marriage made in heaven or earth?* Paper presented at the 24th Lavender Languages & Linguistics Conference, Nottingham, UK.

Arabic and English "Women": A Corpus-Driven Approach to the Politics of the Representation of Women in Online Liberal Discourse

Kamran I. Karimullah
University of Manchester
karimullah.kamran@manchester.ac.uk

The representation of women in the Middle East by European and North American media continues to be a hotly disputed topic. Studies have tied the discursive representation of the women of the Middle East with the political strategies of European interests in the colonial era or the political interests of the United States and its allies after 9/11 (Ahmed 1992, Abu Lughod 2013). Other studies have implicated liberal discourses in service of colonial and contemporary attempts to impose enlightened and liberal reforms on the Middle East in the service of American and European political and economic interests (Massad 2015). These studies highlight how imbalances in global hierarchies of power are manifested at a conceptual level. These studies draw a sharp contrast between the "dominant" European liberal culture and the "receiving" non-European culture.

These studies overstate the element of compulsion and force in the process of communicating liberal thought to the Middle East. For other studies, however, show that the process of communicating liberal ideologies to the Middle East through translation was a process of critical appropriation (e.g. Elshakry 2014, Hourani 1962). These studies also do not attend to the question of how power hierarchies are reflected discursively, at the level of the text itself. Drawing on recent developments in combining CDA and corpus linguistics, I explore the relationship between contemporary online liberal political discourse in Arabic and English. I adopt a corpus-driven approach that has been used by Paul Baker and his collaborators (Baker et al. 2008, 2013a, 2013b). I use an English-language liberal and activist-oriented online sub-corpus (around 500k words) of the Genealogies of Knowledge project (PI Mona Baker, 2016–2019). I construct a corpus of liberal online political Arabic (around 1000k words). As a case study, I generate word sketches for *woman** in each of these corpora. I identify important collocate frames, which I argue index categories of representation of power hierarchies, and the representation of *liberalisms* that are discursively constructed in each corpus as a consequence. Despite the studies by Baker et al. and others, Malamatidou (2018: 63) observes that translation studies is still "lagging behind" in its use of inferential statistics. In addition to employing inferential statistics within the corpora, an important methodological concern is to show that the methods developed by Baker and his collaborators can be used in Arabic corpora, and to what extent they allow us to meaningfully compare patters of representation *between* corpora and in languages with different syntactical properties and rhetorical conventions. Finally, in this study, I shall use the Arabic 2012 corpus (Sketchengine) and English 2013 are comparative corpora to

ensure that the collocates and categories of representation that *woman** attracts are part of a *online liberal political* discourse in each language rather than a consequence of the way people generally tend to discursively represent women online.

I conclude that there is good evidence that the varieties of thought in online Arabic liberal political websites are appropriate to fit local needs. While there is, unfortunately, a truth to the caricatures of an imposed liberalism serving European and American political ends, this analysis shows that liberal discourses are not appropriated uncritically. The numerous frames that form around *woman** in both corpora reveal categories of representation that are rich, many of which are shared, but many others which involve categories that are specific to the representation of women in Arabic- and English-language discourses. I conclude that employing corpus-driven techniques such as word sketches and inferential statistics provide a wholly novel way of approaching debates over the representation of gender and sexuality and power hierarchies in the Middle East, approaches that allow us to compare representation of power in corpora in texts in different languages. Finally, I conclude that using comparative corpora provides reassurance that the representational categories that woman* attracts are a result of a liberal political discourse in Arabic and English, rather than they way tend to speak about women on the internet in English and Arabic.

References

Abu-Lughod, L. 2013. *Do Muslim Women Need Saving*? Cambridge, Mass.: Harvard University Press.

Ahmed, L. 1992. *Women and Gender in Islam*. New Haven: 1992.

Baker, P. et al. 2013a. Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press. Cambridge: Cambridge University Press.

Baker, P. et al. 2013b. "Sketching Muslims: A Corpus Driven Analysis of Representations around the Word 'Muslim' in the British Press 1998–2009". *Applied Linguistics* 34:3, pp. 255–278.

Baker, P. et al. 2008. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press". *Discourse & Society* 19:3, pp. 273–306.

Elshakry, M. 2014. *Reading Darwin in Arabic, 1860–1950*. Chicago: Chicago University Press.

Fairclough, N. 1989. *Language and Power*. Harlow: Longman.

Hourani, A. 1962. *Arabic Thought in the Liberal Age 1789–1939*. Cambridge: Cambridge University Press.

Malamatidou, S. 2018. Corpus Triangulation: Combining Data and Methods in Corpus-Based Translation Studies. London: Routledge.

Massad, J. 2015. *Islam in Liberalism*. Chicago: Chicago University Press.

Childbirth as a political football? A corpus assisted analysis of the discursive construction of "normal birth" in the UK press 1980-2017.

Karen Kinloch
Edge Hill University
kinlochk@edgehill.ac.uk

This paper takes a corpus-assisted discourse analysis of the term "normal birth" in UK broadsheet and tabloid news articles from 1980 to the present day (2017), with particular reference to diachronic differences and similarities in the discursive construction of "normal birth". This study is most pertinent in light of recent reporting in the UK press of the change in terminology by the Royal College of Midwives (which was in fact instigated in 2014) from "normal" to "physiological" birth. To hail this change in language as a victory for choice elides

the problematic dichotomy between so called medical and midwifery models (Downe, 2011), and the conflict between the proponents of these fields around what constitutes normalcy in childbirth.

Drawing on previous work combining corpus linguistics and discourse analysis (Baker, 2006; Baker et al., 2008; Partington, 2008), applied to news reports on a socially problematic or contested term, this study focuses on keywords and collocates elicited from a specialist corpus of news texts which include the term "normal birth". This specialist corpus was compiled of UK national news texts including the search terms "normal birth" or "normal births" retrieved from the Nexis database including all time periods which are currently available via this resource. This resulted in a corpus of 302 texts distributed as shown in Table 1.

Decade	Texts	Tokens
1980s	6	6,349
1990s	32	22,589
2000s	151	129,296
2010s	113	98,452
Total	302	256,686

Table 1. Total texts and tokens in the "normal birth" corpus by decade.

Particularly relevant to the methodology of this paper is diachronic approach utilised in Corpus-Assisted Discourse Studies (CADS) (Partington, 2008) of news texts in which keywords for a particular time period are compared in order to elicit key differences and similarities in the lexis of texts produced in this time.

Using SketchEngine (Kilgariff et al., 2004) to generate keywords lists, a list of the top 50 keywords for each decade were created and then thematically coded into categories which include; individual actors, named institutions, experiences, processes, locations and body parts. This paper is focused on the category of actor keywords, which in this data set encompassed midwives, obstetricians, gynaecologists and "mothers-to-be" and the representation of these actors both individually and collectively.

WordSketches (Kilgariff et al., 2004) showing collocates of these "key" actors were used to illuminate the lexis used to represent them in news discourse on "normal birth". These named actors are positioned in discourses of risk, and financial and physical costs and, whilst there were topical differences in the reporting of individual cases, there was little change in the contested and problematic nature of this term across the time period studied.

It became clear that the representations in these texts reinforced the positioning of obstetricians and midwives in conflict, both resisting the validity of the others position in the dialogue of what "normal" birth can be. The reinforcement of this dichotomy in the public domain potentially serves to limit women's choices around birth and background their voices.

References

Baker, P. 2006. *Using corpora in discourse analysis.* London, Continuum.

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T. and Wodak, R. 2008. "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press". *Discourse & Society*, 19(3), pp.273-306.

Downe, S. 2011. *Normal childbirth.* Edinburgh: Churchill Livingstone.

Kilgarriff, A., Rychlý, P., Smrz, P. and Tugwell, D. 2004. The Sketch Engine. *Proc EURALEX 200.* Lorient, France, 105–116.

Partington, A. 2008. "The armchair and the machine: Corpus-Assisted Discourse Studies" in C. Taylor Torsello, K. Ackerley and E. Castello (eds.) *Corpora for University Language Teachers.* Bern: Peter Lang. 189–213

Interdisciplinarity and CADS: Critical reflections from a case study of online groomers' language

Anina Kinzel
Swansea University
795737@swansea.ac.uk

This poster presentation will reflect critically upon the scientific concept of interdisciplinarity by examining how two methodologies, underpinned by different Social Sciences disciplines, approach the same research question and data set. Specifically, I will consider how Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2015) – within Psychology - and CADS – within Linguistics (Partington, 2008; Hardt-Mautner, 1995) - may be used to investigate online groomers' language in a large dataset of chat logs taken from the Perverted Justice website.[2] This study is part of a research partnership project between Swansea University and the NSPCC.[3]

Although CADS is broadly deemed an interdisciplinary methodology (e.g. Partington, 2007; Baker & McEnery, 2015), instances of interdisciplinary work are less frequent than might be expected. As Marchi et al. (2017: 174) put it: 'Because of its explorative nature and omnivorous interests, CADS appears to have an intrinsic interdisciplinary vocation, and yet there are very few examples of actual interdisciplinary research to date.' This may be partly linked to considerable debate in CADS - and more broadly in Linguistics and other disciplines – regarding what counts – or not – as interdisciplinary research.

Some scholars consider research that is informed by more than one discipline to be interdisciplinary (Frey, 2004), while others argue that only research that *integrates* methods from two or more separate disciplines is truly interdisciplinary (Marchi et al., 2017; Lorenzo-Dus & Marsh, 2012). Van Leeuwen (2005) proposes three different models of interdisciplinarity: the centralist, pluralist and integrationist models. The centralist model

---

[2] The Perverted Justice Foundation is a non-profit organisation that uses on volunteers pretending to be teenagers entering chat logs and waiting to be approached by online groomers. They then collaborate with law enforcement to convict the groomers, at which point the resulting chat log is uploaded to their publicly available database. (http://www.perverted-justice.com/)

[3] Project title: '*Online Grooming Communication: From Research to Prevention'; project funded by the EPSRC (CHERISH-DE Centre). Principal Investigator:* Prof N Lorenzo-Dus (Swansea University).

considers a discipline to be the 'centre of the universe of knowledge' (van Leeuwen, 2005: 3) merely touching upon other disciplines e.g. to provide context to an analysis. The pluralist model emphasises problems that may be addressed by more than one discipline, seeking to combine these disciplines as 'equal partners' (van Leeuwen, 2005: 6). The integrationist model also highlights problems that may be addressed by more than one discipline. However, it acknowledges that any given discipline cannot address a problem independently. Instead, disciplines are regarded as interdependent, which calls into question the definition of disciplines and their boundaries.

Taking the above into consideration, and drawing upon data from online grooming chat logs – I critically (1) appraise the advantages and disadvantages of using LIWC and CADS for identifying online groomers' communicative patterns; and (2) explore the extent to which these methodologies may be integrated as 'equal partners' in common efforts to contribute to prevent online grooming cases. My work suggests that there are many advantages to drawing on both Psychology and Linguistics for this type of research. It also shows that a CADS approach enhances an analysis that uses LIWC and, therefore, that the two methodologies can be integrated effectively. In terms of CADS, disciplinary boundaries need to be tested and reviewed and progressively a move towards a more integrationist model of interdisciplinary is needed to fulfil CADS' applied research agenda.

References

Baker, P. and McEnery, T. 2015. "Introduction". In T. McEnery and P. Baker (eds.). *Corpora and Discourse Studies: Integrating Discourse and Corpora*. New York: Palgrave.

Frey, J. 2004. "The Cultural Shaping of ICTs within Academic Fields: Corpus-based Linguistics as a Case Study". *Literary and Linguistic Computing*. 19 (3). 303-319.

Hardt-Mautner, G. 1995. "'Only Connect.'- Critical Discourse Analysis and Corpus Linguistics. UCREL Technical Paper 6". Lancaster: University of Lancaster. Available from: http://stig.lancs.ac.uk/papers/techpaper/vol6.pdf (last accessed January 2018).

Lorenzo-Dus, N. and Marsh, S. 2012. "Bridging the gap: Interdisciplinary insights into the securitization of poverty". *Discourse & Society*. 23 (3). 274–296.

Marchi, A., Lorenzo-Dus, N. and Marsh, S. 2017. "Churchill's inter-subjective special relationship – A corpus assisted discourse approach". In A. P. Dobson and S. Marsh. (eds.) *Churchill and the Anglo-American Special Relationship*. New York: Routledge.

Partington, A. 2007. "Metaphors, motifs and similes across discourse types: Corpus-Assisted Discourse Studies (CADS) at work". In A. Stefanowitsch and S. T. Gries (eds.) *Corpus-based approaches to metaphor and metonymy* (Vol. 171). Berlin and New York: Walter de Gruyter.

Partington, A. 2008. "The armchair and the machine: Corpus-Assisted Discourse Studies". In C. Taylor Torsello, K. Ackerley and E. Castello (eds). *Corpora for University Language Teachers*. Bern: Peter Lang.

Pennebaker, J.W., Booth, R.J., Boyd, R.L., and Francis, M.E. 2015. "Linguistic Inquiry and Word Count: LIWC2015." [Software]. Austin, TX: Pennebaker Conglomerates. Available from: www.LIWC.net (last accessed January 2018).

Van Leeuwen, T. 2005. "Three models of interdisciplinarity". In R. Wodak and P. A. Chilton (eds.) *A New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity*. Amsterdam: Benjamins.

Multimodal corpus analysis of representations of travel destinations: two methodological approaches

Ekaterina Ignatova
Lancaster University
e.ignatova@lancaster.ac.uk

This poster presentation introduces two approaches to multimodal corpus analysis aimed at comparing how two travel destinations, Moscow and London, are represented in tourism discourse of a popular travel website www.tripadvisor.com.

In recent years there has been a surge of interest to utilizing multimodal approaches, which consider various expressive resources used for making meaning in texts, for instance, writing, images, layout (Van Leeuwen, 2015), for analysing tourism discourse (e.g. Francesconi, 2011, 2014). However, methodological frameworks for studying the representation of travel destinations in multimodal texts have not received much attention. The project aims at filling this gap by discussing two multimodal approaches to city representation analysis in travel-related texts.

Due to the fact that such modes as layout, typography and hypertextual structure are predetermined by the website design and are similar for both cities, the primary concern of the current project is language and images.

In both described approaches, corpus linguistics techniques, such as keywords comparison, concordance analysis of keywords, collocation comparison and concordance analysis of collocations, are applied for language analysis in order to search for the salient features of the texts and to compare how the two cities are represented.

As for visual analysis, the first option implies a qualitative study of images, which is a popular approach in the field of tourism discourse. In addition to commonly used social semiotics visual analysis (Kress & Van Leeuwen, 2006), the proposed methodology includes a comparative and interpretative analysis based on Dann's (1996) framework of visual techniques and Barthes' (1977) theory of visual semiotics. The advantage of this methodology is that it allows to get deeper insights into what visual elements, structures and techniques are utilized to construct the representation of Moscow and London in the photographs on the website. The drawbacks, however, are that such analysis is time-consuming, consequently, only a limited number of images can be studied.

The second approach combines the aforementioned corpus linguistics methods with an analogous visual mode analysis based on tagging images with topic words and comparing frequencies of the topic words across the corpora. This stage is aimed at identifying patterns in the use of images for representing travel destinations and comparing them between the two corpora. In addition, the analysis of how the images are used in context, namely, which texts they accompany, is proposed to interpret the patterns. This methodology can be applied to a larger number of multimodal texts and allows the exploration of similarities and differences in the representation of the two cities. Therefore, the two approaches complement each other and can be used individually or in combination depending on the time limit and research aims.

The project is a further step in the development of multimodal corpus approaches to the study of the representation of travel destinations in tourism discourse.

References

Barthes, R. 1977. *Image Music Text*. London: FontanaPress.
Dann, G. 1996. *The Language of Tourism: A Sociolinguistic Perspective*. Oxon: CAB INTERNATIONAL.
Francesconi, S. 2011. Images and writing in tourist brochures. *Journal of Tourism and Cultural Change*, 9(4), 341-356.

Francesconi, S. 2014. *Reading tourism texts: a multimodal analysis* (Vol. 36): Channel view publications.

Kress, G., & Van Leeuwen, T. 2006. *Reading Images: The Grammar of Visual Design* (2nd ed.). London, New York: Routledge.

Van Leeuwen, T. 2015. Multimodality. In D. Tannen, H. Hamilton, & S. Deborah (Eds.), *The Handbook of Discourse Analysis* (2nd ed., pp. 447-465): John Wiley & Sons, Inc.

Making Hong Kong a smarter travel destination: A corpus-assisted discourse study of travel forum posts on social media

Phoenix W. Y. Lam
The Hong Kong Polytechnic University
wyplam@polyu.edu.hk

In the digital age, travellers are increasingly doing research online in preparation for their trips. One major platform for doing such online research is through travel forums, which have been playing a significant role in knowledge sharing in the travel planning process (Okazaki et al. 2017). However, surprisingly little attention has been paid to these online travel forums in discourse studies. In these user-generated interactive forums for travel destinations around the world, users are free to post any message or reply to message posted by other users. A typical destination forum can attract hundreds of thousands of messages on virtually any topic deemed to be relevant to the destination. Such forums are therefore closely related but functionally different from the highly evaluative online consumer reviews on service items such as hotels and restaurants, which also represent examples of online travel discourse often found on the same travel websites but have received far more scholarly attention thus far in discourse studies (see, for example, Vasquez 2011; Ho 2017).

With the aim of better understanding the nature of the scarcely researched online travel forums, the present study examines the Hong Kong travel forum on the popular travel website TripAdvisor. As the first large-scale corpus-assisted discourse analytic study of travel forum posts, it investigates more than 2,500 posts submitted to the forum in a year. Specifically, the three research questions guiding the present study are: 1) What are the key functions of the travel forum posts and are there prominent features associated with the functions identified? 2) What are the main topics of interest regarding Hong Kong as revealed by the forum posts? 3) Based on the findings from 1) and 2), what specific suggestions can be made to make Hong Kong a smarter travel destination?

To answer the three research questions listed above, all the posts collected were first read and analysed closely to identify their functions and themes manually in a detailed content analysis. Once they were classified both functionally and thematically, the posts serving the key functions or expressing the main themes were examined in groups in more detail, both qualitatively using discourse-analytic techniques and quantitatively by corpus-linguistic methods such as the study of the most frequently-occurring items and keywords. Unlike most corpus-assisted discourse studies which often start with a frequency list to identify areas of interest for further examination, the present study first proceeded with the manual coding of the posts to identify the most salient functions and themes, followed by a more focused frequency-based analysis of the items associated with the dominant functions and themes. It is believed that this approach can reveal distributional patterns which are more directly related to specific functions and themes, which in turn will provide a more nuanced understanding of the nature of

the online travel forum concerned.

Findings from the present study show that most of the forum posts are informational in function, either seeking information from or giving information to other forum users. This confirms that travel forum posts are functionally distinct from the evaluative travel product reviews, involving different discursive practices and levels of interactivity. The main topics of interest regarding Hong Kong as revealed by the forum posts are transportation, attractions and activities. The closer analysis of these main topics into sub-topics yields further insights into the major areas of concern by travellers worldwide, which allow evidence-based concrete recommendations to be made on tourism policies and strategies. Importantly then, the present study demonstrates the value of a corpus-assisted discourse study in addressing real-world issues in the domain of tourism.

References

Ho, V. 2017. Achieving service recovery through responding to negative online reviews. *Discourse & Communication* 11(1): 31-50.

Okazaki, S. Andreu, L. & Campo, S. 2017. Knowledge sharing among tourists via social media: A comparison between Facebook and TripAdvisor. *International Journal of Tourism Research* 19: 107-119.

Vásquez, C. 2011. Complaints online: The case of TripAdvisor. *Journal of Pragmatics* 43(6): 1707-1717.

A Corpus-based Analysis of North Korean Defectors in Public Discourse

Sun-Hee Lee
Wellesley College Yonsei University
slee6@wellesley.edu

Beomil Kang
Wellesley College Yonsei University
kangbeomil@gmail.com

In March 2017, the number of North Korean Defectors (NKDs) residing in South Korea reached 30,490 according to the Ministry of Unification. Despite a surge in the number of defectors over two decades, socio-economic challenges and prejudice against those who crossed the border continues to intensify. In the wake of fluctuating tensions with North Korea's nuclear threats and a new administrative agenda for the South Korean government, the public attention to NKDs seems to be significantly diluted. While the recent announcement of North Korea's participation in the Pyeongchang Winter Olympics promotes expectations for a release of tensions in the Korean Peninsula, *NAVER*, the largest Korean news portal carried some articles of NKDs protesting against the North's Olympics participation. Interestingly those headlines address the NKDs using the term *Thalbukja-teul* (plural), literally meaning 'the people who escaped from the North' instead of *Saytheomin* (both singular and plural) meaning 'the new settlers', which has been promoted by the Ministry of Unification since 2005. In addition, the nominal *Thalbukjateul* combines with a modifying expression *Hankuk nay* 'within South Korea' instead of the common adjective *ceonkuk* 'nationwide'. The linguistic expression *Hankuk nay* implies exclusion of NKDs from the citizenship of South Korea while categorizing their status as migrants. This example demonstrates how language functions as a significant tool for representing or dividing identity, status, and power among individuals and nations. In this paper, we provide a data-driven critical discourse analysis of public discourse regarding NKDs. The analysis examines how media function to formulate the identity of NKDs and stereotypes/prejudices through linguistic representation.

Critical discourse analysis (CDA) defines language as a social practice and examines how ideologies and power relations are represented in language. Noting how power is exercised through language in social and political structure, this study aims to provide a corpus-based discourse analysis on how the South Korean news media identify, categorize, and represent NKDs and to explore the dynamics of language, identity, and power in public discourse. The analysis of public discourse must be interrogated from broad realms of social, historical, and political contexts. In conjunction with the long-term research project on the comprehensive discourse analysis of NKDs, our current work focuses on four major broadsheet newspapers that have distinct political stances and investigates interactive discourse features that contribute to representations of NKDs in the South Korean community.

In line with previous work of Baker, Gabrielatos and McEnery (2013) on British media attitudes toward Muslims post 9/11, this study elicits empirical linguistic patterns and distributions in conjunction with media attitudes and representations. Methodologically we combine methods of corpus-based analysis and critical discourse analysis. The corpus data were built by collecting over 29,566 news texts on defectors from the *Chosun Ilbo*, *Dong-A Ilbo*, the *Hankyore,* and the *Kyunghyang* published between 1994 and 2017. Texts have been annotated and further processed to extract relevant linguistic information. Linguistic distributions of topic words, key words, and collocations are explored and classified using statistical tools and semantic and pragmatic analysis, as well as discourse analysis. Both quantitative and qualitative linguistic analysis substantiates similarities and differences in the ways that South Korean newspapers categorize defectors and address related issues.

Our findings indicate that the South Korean news media are mostly focusing on domestic or international political issues rather than issues of individuals or human rights. Overall, the predominant usage of words with strong negative discourse prosody is distinct across all news texts. We conclude that the presentation of NKDs in South Korean newspapers between 1994 and 2017 was carried out in negative contexts related to political conflicts with North Korea while empirical issues of the NKDs' settlement in the community were barely addressed. The analysis of topic words and collocations indicates that there are more issues with respect to groups of women, youths, and orphans. We further examine variations of individual newspapers and changes in time across the corpus while exploring different political stances and concerns between different newspapers and diachronic change of keywords and discourse prosody of conflicts. Both qualitative analysis and quantitative tools will explicate how language is used to substantiate stereotypes and bias by media, which result in a credible analysis. The outcomes are expected to reveal empirical issues and challenges not only for current South Korean society and its inclusion of NKDs but also for a reunified Korea in the future.

References

Baker. P. and McEnery, T. 2005. "A Corpus-Based Approach to Discourses of Refugees and Asylum Seekers in UN and Newspaper Texts". *Journal of Language and Politics*.4 (2).197-226.

Baker, P. Gabrielatos, C. and McEnery, T. 2013. *Discourse Analysis and Media Attitudes*, Cambridge University Press, Cambridge.

Fairclough, N. 1989. *Language and Power*. London: Longman.

Ministry of Reunification. 2015. *2014 Report on Current Situations of North Korean Defectors*. Government of Republic of Korea.

Skuttnab-Kangas, T. and Phillipson, R. 1994. *Linguistic Human Rights*. Berlin: Mouton de Gruyter,

Creating a corpus of social workers' writing: methodological challenges, representational issues and analytical concerns

Maria Leedham
The Open University
Maria.leedham@open.ac.uk

Theresa Lillis
The Open University
Theresa.lillis@open.ac.uk

Alison Twiner
The Open University
alison.twiner@open.ac.uk

Within the profession of social work, the production and use of written texts is a high-stakes activity, playing a central role in all decisions around providing services for people, and at the same time used to evaluate social workers' professional competence. Writing of all kinds pervades everyday social work practice, varying from case notes to inter-departmental emails to formal assessment reports. Social work writing (often subsumed under the label of 'recording') is frequently the target of criticism in formal reviews and public media reporting of social work practice, hitting headline news when a case of extreme abuse or death occurs. Despite this, little empirical research has been carried out on these texts and, in particular, no corpus of professional UK social work writing exists. This gap is addressed by the WiSP project ('Writing in professional social work practice in a changing communicative landscape') which brings corpus work together with discourse analysis in an ethnographically-framed, 2.5-year project. The specific focus of this paper is on the corpus linguistic dimension of the research, and in particular the methodological, representational and analytical challenges and opportunities this entails.

The newly-created WiSP corpus comprises 1 million words within 4,600 texts from the writing of 38 social workers within three UK local authorities in 2015-2017, and covers the domains of adult generic, adult mental health and children's care. Insights from corpus analysis are interpreted with the understandings gained from observations (10 social worker weeks), interviews (with 70 practising social workers across five local authorities) and additional documentary evidence (see Lillis, Leedham and Twiner, in progress).

In this paper, we describe the value of building a corpus of hard-to-reach social work texts, the methodological difficulties around gaining access to the texts due to their very sensitive nature and the challenges inherent in on-site text anonymization. We consider how the difficult-to-access nature of the texts meant that we were less able to configure corpus design in advance and were less in control of the range, granularity and consistency of anonymization codes. This perhaps undesirable research situation has, however, offered valuable insights into the terms used by insider social work staff, for instance between the use of 'service user', 'client' or 'customer'.

The paper will feature findings from keyword analysis of the corpus overall and of the subcorpora of different text types, using Wmatrix (Rayson, 2008) software to extract key items and sort on the measure of effect size (Gabrielatos, 2018), combined with thematic categorisation using concordance lines, collocates, and close reading for greater context. The use of corpus based discourse analysis illustrates the prevalence of *people* within social work; this is highlighted by the aggregation of all personal names through the various anonymization codes. Thus family and institutionalised care is a prominent category, as is the category of related professional services such as medical, mental health, educational and legal support. A further finding has been around the different ways social workers report in their writing on their own or colleagues' communications and actions and how they report the communications and

actions of the people they are supporting, in terms of levels of formality and personalisation.

References

Gabrielatos, C., 2018. "Keyness Analysis: nature, metrics and techniques". In C. Taylor and A. Marchi (eds.) *Corpus Approaches to Discourse: A critical review.* Chapter 12. Oxford: Routledge.

Lillis, T., Leedham, M. and Twiner, A. (in progress) "'If it's not written down it didn't happen': contemporary writing and recording demands in social work". For *Journal of Applied Linguistics and Professional Practice.*

Rayson, P. 2008. "From key words to key semantic domains". *International Journal of Corpus Linguistics.* 13:4 pp. 519-549. DOI: 10.1075/ijcl.13.4.06ray

A Study of Customer Support Dialog Chat Conversations

Jochen L. Leidner
Thomson Reuters, Research & Development,
leidner@acm.org

Customer support functions provide a set of customer support agents that help customers who have questions about products or who face technical or business issues. Nowadays, synchronous online chat is one of the technical channels through which an information-seeking or help-seeking customer can interact with a support agent.

The present poster paper is a study of a real corpus of product usage related customer chat by financial professionals. Unlike artificial collaborative problem solving (e.g. in the MAPTASK project, cf. Anderson *et al,* (1991)), customers often face a pressure to solve their problem in a limited time. We provide some descriptive corpus statistics and investigate the discourse structure of typical turn sequences and attempt a classification of them. We also provide a discussion how collaborative problem-solving influences the conversation compared to more entertaining related chats studied in the past. We also discuss the ethical issues including privacy around conducting work on real chat. While the analysis of online chat and forum posts have received some attention (Cherny (1999), Schulze (1999), Ooi (2002), Goutsos (2005), Forsyth (2007), Shaoul & Westbury (2010), Uthus & Aha (2013a,b,c), *inter alia*), we do not know of any previous work on real customer chats.

References

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G. M., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S. and Weinert, R. (1991). "The HCRC Map Task Corpus" *Language and Speech*, **34**, pp. 351-366.

Cherny, L. (1999). *Conversation and community: Chat in a virtual world.* Stanford, CA: CSLI Publications.

Forsyth, Eric N. (2007). "Lexical and Discourse Analysis of Online Chat Dialog" http://hdl.handle.net/10945/39587

Forsyth, Richard, David Clarke and Phoenix Lam (2008). "Timelines, talk and transcription: A chronometric approach to simultaneous speech" *International Journal of Corpus Linguistics* 13*(2)*, 225-250.

Goutsos, D. (2005). "The interaction of generic structure and inter- personal relations in two-party e-chat discourse." *language@internet* **2**, article 3. (urn:nbn:de:0009-7-1885)

MAPTASK http://groups.inf.ed.ac.uk/maptask/ (cited 2018-01-31)

Ooi (2002) "Aspects of computer-mediated communications for research in corpus linguistics", in: New Frontiers of Corpus Research: Papers from the Twenty First International, ed. by Pam Peters, Peter Collins, Adam Smith.

Shaoul, C. & Westbury C. (2010). A USENET corpus (2005-2010) Edmonton, AB: University Alberta,

Unpublished data set.

Schulze, Markus (1999). "Substitution of paraverbal and nonverbal cues in the written medium of IRC, in: Naumann, Bernd (ed.) *Dialogue Analysis and the Mass Media* Tubingen: Niemeyer.

Uthus, David C. & Aha, David W. (2013a). "The Ubuntu Chat Corpus for Multiparticipant Chat Analysis." Papers from the 2013 *AAAI Spring Symposium on Analyzing Microtext*, 99-102. Stanford, CA: AAAI Press

Uthus, David C. & Aha, David W. (2013b). "Detecting Bot-Answerable Questions in Ubuntu Chat" *International Joint Conference of Natural Language Processing*, Nagoya, Japan, 14-18 October 2013, 747-752.

Uthus, David C., & Aha, David W. (2013c). "Multiparticipant Chat Analysis: A Survey" *Artificial Intelligence*. 199-201: 106-201, doi:10.1016/j.artint.2013.02.004

No means no is not enough: representations of consent in erotic writing

Alon Lischinsky
Oxford Brookes University
alon@lischinsky.net

Kat Gupta
University of Sussex
kat@mixosaurus.co.uk

Recent public debates about the 'epidemic of sexual harassment' in politics, the entertainment industry and the workplace more generally have drawn considerable attention to the role of pornography in shaping cultural ideas about sexuality and gender relations. Critical voices have argued that porn fosters the misogynistic belief that 'the stealing and buying and selling of women are not acts of force or abuse because women want to be raped and prostituted because that is the nature of women and the nature of women sexuality' (Dworkin 1993, 240); the stories of insatiable female desire hidden behind a token resistance that (are claimed to) characterise porn encourage men to adopt an aggressive stance in the negotiation of sexual encounters, and to coerce prospective partners into sex rather than seeking their consent.

The notion of consent seems, however, more complex than these debates tend to acknowledge. Radical feminist criticism typically takes as their starting point a totalising and monolithic view of patriarchal structures, in which the possibilities for female agency are entirely occluded (Paasonen 2011, 54). Under such conditions consent would be ex hypothesi impossible: having been thoroughly socialised into the belief that their bodies are meant for male pleasure, women lack the freedom necessary to meaningfully communicate their own desires and their limits. Conversely, psychological and legal approaches examine the practices involved in the negotiation of sexual encounters from the assumption that all participants are on equal terms, and consequently '[fail] to account for wider social inequalities, […] external pressures or power structures' that may shape such choices (Popova 2017)— such as the stigma attached to female promiscuity, or the frequent belief that female refusals are a face-saving token of resistance rather than a genuine expression of dissent.

Crucially, both of these views derive their understanding of consent from a pre-existing and more abstract theory of social relations, rather than constructing it bottom-up from an examination of the actual discursive practices that participants use to negotiate and make sense of sexual encounters. Analyses taking this latter approach have yielded a much more nuanced view of culturally-normative sexual scripts, showing that both males and females have a sophisticated understanding of the acceptable ways of performing refusal, yet are able to strategically use claims of misunderstanding as 'self-interested justifications for coercive sexual behaviour' (O'Bryne et al 2006, 135). Examination of such scripts can therefore provide insights

into cultural notions of consent that avoid both the Scylla of total structure and the Charybdis of unconstrained agency.

Knowledge of such scripts —the repertoire of rules and expectations governing the expression, interpretation and performance of sexual activity— is not solely or even primarily acquired through direct experience (Laumann and Gagnon 1995). Sexuality is one of many domains of social life where first-hand knowledge is limited because of the privacy, stigma and taboo that surrounds it, and individuals rarely have the opportunity to witness a sufficient range of encounters to form a clear idea of the range of allowable behaviours. In such cases, it is almost impossible to disentangle the object itself from the skein of cultural narratives and imaginings that preform it: sex is not only experienced first but also more frequently and with greater variety through fiction than in real life (author, in press), and pornography is the primary source of such experiences. Even though porn audiences show a clear awareness of the distinction between pornographic fantasy and real-life sexual interaction (McKee 2010), this distinction is not always reflected in sense-making, and beliefs and attitudes derived from fictional media can blend with non-fictional ones in their general knowledge (Marsh et al. 2003). This suggests that the typical scripts depicted in porn may exert considerable influence in shaping audiences' expectations for sexual encounters, including their negotiation (Popova 2015), especially as alternate sources of information —for example in sex education materials— are often lacking or inadequate (Albury 2014).

Feminist porn auteurs such as Tristan Taormino have explicitly embraced this educational function (Voss 2014), giving communication and consent a salient role in their works, and the implications of such alternative pornographies have been extensively debated (cf. Taormino et al. 2013). However, examinations of notion of consent conveyed (largely in an inexplicit manner) in more mainstream pornographic materials remain rare. This project seeks to fill this gap by examining the representations of sexual negotiation, refusal and coercion in a large body of amateur-authored erotic narratives drawn from Literotica.com, one of the oldest and largest erotic fiction repositories online. Drawing on the seminal work of Popova (2015, 2017) as well as the conversation-analytic literature on sexual refusals (Kitzinger & Frith 1999; O'Byrne et al. 2006), we seek to: first, identify the range of textual signals associated with the requesting, granting, refusing or withdrawing of consent, including characters' explicit statements, nonverbal cues such as gesture or facial expression, and indirect cues provided by narrative structure, focalisation and mind style (Semino 2007); and second, examine their prevalence across the corpus and their role in the stories' overall narrative structure.

References

Albury, K. (2014). Porn and sex education, porn as sex education. *Porn Studies*, 1(1–2), 172–181.

Attwood, F. (2011). Sex and the media. In Ross, K. (Ed.) *The Handbook of Gender, Sex, and Media* (pp. 455–469). Oxford: Wiley-Blackwell.

Dworkin, A. (1993). Letters from a War Zone. Chicago, IL: Lawrence Hill.

Kitzinger, C., & Frith, H. (1999). Just say no? The use of conversation analysis in developing a feminist perspective on sexual refusal. *Discourse & Society*, 10(3), 293-316.

Laumann, E. O. & Gagnon, J. H. (1995). *A sociological perspective on sexual action*. In Parker, R. G. & Gagnon, J. H. (Eds.). Conceiving Sexuality: Approaches to Sex Research in a Postmodern World, (pp. 183–214). New York: Routledge.

Marsh, E. J., Meade, M. L., & Roediger III, H. L. 2003. Learning facts from fiction. *Journal of Memory and Language*, 49(4), 519–536.

McKee, A. (2010). Does pornography harm young people? *Australian Journal of Communication*, 37(1), 17–36.

McNair, B. (2014). Rethinking the effects paradigm in porn studies. *Porn Studies*, 1(1–2), 161–171.

O'Byrne, R., Rapley, M., & Hansen, S. (2006). 'You couldn't say "no", could you?': Young men's understandings of sexual refusal. *Feminism & Psychology*, *16*(2), 133-154.

Paasonen, S. (2011). *Carnal Resonance. Affect and Online Pornography*. Cambridge, MA: MIT Press.

Popova, M. (2015). Researching consent: How popular and vernacular depictions of consent can extend academic understanding. Paper presented to the *Researching Sex and Sexualities* conference, May 2015

Popova, M. (2017). 'Dogfuck Rapeworld': Omegaverse fanfiction as a critical tool in analysing the impact of social power structures on intimate relationships and sexual consent. *Porn Studies*, published online ahead of print.

Semino, E. (2007). Mind style twenty-five years on. *Style*, *41*(2), 153-172.

Taormino, T., Penley, C., Shimizu, C., & Miller-Young, M. (Eds.). (2013). *The Feminist Porn Book: The Politics of Producing Pleasure*. New York: The Feminist Press at CUNY.

Voss, G. (2014). Tristan Taormino interviewed by Georgina Voss. *Porn Studies*, *1*(1–2), 203-205.

Modifying nouns, depersonalizing medicine: A corpus-based analysis of noun phrase complexity in medical prose from 1700 to the present and its implications for the sociocultural history of medicine

Georg Marko
Karl-Franzens-University Graz
georg.marko@uni-graz.at

One of the characteristic features of today's medical prose[4] – and of scientific registers in general – is high noun phrase complexity:[5] the meaning of a noun is specified or expanded by adding information in nominal compounds, e.g. *areas* → *skin areas*, in premodifying adjectives, e.g. → *sensitive skin areas*, in postmodifying prepositional phrases, e.g. → *sensitive skin areas of their bodies*, or in relative clauses, e.g. → *sensitive skin areas of their bodies that are not covered*, to name just a few relevant structures. According to Biber & Conrad (2009: 167-168), this has not always been the case, but NPs have become increasingly complex in the history of medical writing in English over the past three centuries. Some constructions – especially nominal compounds and premodifying adjectives – have contributed to this change more significantly than others – especially relative clauses, the former being semantically 'compressed', i.e. *less* explicit, and the latter 'expanded', i.e. *more* explicit (Biber & Clark 2002: 63).

The frequent use of highly complex noun phrases, particularly of the more 'compressed' kind, may lead to conceptual condensation (being able to pack more content into 'less text'), technicalization (full comprehension requiring specialist knowledge and expertise), and – with the increased emphasis on nouns – reification (conceptualizing the world in terms of entities rather than processes). These aspects can all be associated with depersonalization (backgrounding the importance of patients, their understanding and experience of a situation). The diachronic change of complexity could thus be argued to reflect and possibly also promote the shift of medicine from a patient-centred to a technical discipline more interested in abstract conditions than in concrete persons over the last 300 years (Jewson 1976, Armstrong 1995).

My study uses corpus linguistic methods to examine noun phrase complexity and the different types of noun phrase modifications and trace their developments in a diachronic corpus of

---

[4] The present study focuses on English, but results would probably be similar for other languages.

[5] For reasons of simplicity, this also encompasses the complexity of nouns themselves, e.g. in nominal compounds.

medical prose. It builds on Biber's (et al.) work on the subject, trying to replicate his results. But at the same time the study expands his approach, looking at more structural categories in a larger corpus with more sociocultural contextualization. The main goal is to gain a detailed and differentiated picture concerning which structures increased in frequency at which times, how strongly, and whether linguistic results correspond to developments in the sociocultural history of medicine.

The study combines the corpus-based versions of Critical Discourse Analysis (e.g. Mautner 2009) and diachronic register analysis (e.g. Biber & Conrad 2009). The corpus used is a self-compiled collection of medical articles covering the period from 1700 to the present (one article for each year). The texts have been taken from (mainly British) general medical journals, e.g. *Philosophical Transactions of the Royal Society*, *Edinburgh Medical Journal*, and *The Lancet*. The corpus comprises approximately 850.000 words.

References

Armstrong, D. 1995. "The rise of surveillance medicine". *Sociology of Health and Illness* 17(3). 393-404.

Biber, D. & S. Conrad 2009. *Register, genre and style*. Cambridge: CUP.

Biber, D. & V. Clark 2002. "Historical shifts in modification patterns with complex noun phrase structures. How long can you go without a verb?" In T. Fanego et al. (eds.). *English historical syntax and morphology*. Amsterdam: Benjamins. 43-66.

Jewson, N. 1976. "The Disappearance of the Sick Man from Medical Cosmology 1770-1870." *Sociology* 10. 225-244.

Mautner, Gerlinde 2009. "Checks and Balances: How Corpus Linguistics can Contribute to CDA". In R. Wodak & M. Meyer (eds.). *Methods of Critical Discourse Analysis*. 2nd edition. Los Angeles: Sage. 122-143.

What your followers say about you: a dialectical-relational approach to (collective) identity in the followership of an online protest movement

Mark McGlashan
Birmingham City University
Mark.McGlashan@bcu.ac.uk

This paper investigates (collective) identity/identities and methods of investigating an online protest movement as a site for ambient affiliation, "where […] individuals do not necessarily have to interact directly, but may engage in mass practices such as hashtagging in order to participate in particular kinds of 'belonging'" (Zappavigna 2017: 216). More specifically, this paper focuses on the Football Lads Alliance (FLA) Twitter account and its followers and examines the dialectical relationship between how the FLA defines itself and the self-reported identities evident in the biographical data of its online followers. The paper will first introduce and contextualise the FLA before exploring methods of data collection/analysis and presenting findings from the research.

The FLA was founded on 4th June 2017 – the day after the *London Bridge Attack* – by John Meighan who has a conviction for football-related violence. The group exists and operates mainly online using platforms like Facebook, Twitter and YouTube but has also organised and

held two marches through central London. The first being on 24th June 2017 at London Bridge, the site of the *London Bridge Attack*, and the second on 7th October 2017. The FLA identifies as an anti-extremism protest movement, the focus of which is briefly summarised in the group's Twitter[6] biography as "Uniting the Football Family Against Extremism", a sentiment also found in the FLA facebook group[7] 'About' page, which says, "We are a new movement with a new purpose to fight extremism, we in no way condone racist behaviours. Enough is enough, it is time to stand up. Disclaimer – Comments made on our posts are not monitored and DO NOT represent the views of the FLA." However, the group's aims and ambitions are evolving as is suggested by a more detailed manifesto given on FLA's website[8], which outlines a range of beliefs ("a safer environment […] for all of our children and grandchildren", cultural inclusivity, political accountability) and aims ("greater restrictions placed on known terror suspects", review of terrorism legislation, community cohesion, greater support for veterans and victims of terrorism). Despite the group's numerous explicit claims of cultural inclusivity and rejection of racism, sexism, bigotry, etc., some of the FLA's behaviours suggest discursive links with individuals and groups that do espouse extremist ideological positions (specifically, islamophobia). Toni Bugle[9], founder of Mothers Against Radical Islam and Sharia (MARIAs) and member of the nationalist English Democrats party, spoke at the first FLA march. Meighan has also given one of his few interviews to the blog Shy Society, which features posts on topics like race, free speech, and membership of the European Union, and has also interviewed Anne Marie Waters[10], leader of For Britain, a far-right offshoot of the United Kingdom Independence Party (UKIP), founding member of Sharia Watch UK[11], and co-leader of the anti-Islam Pegida UK[12].

This paper understands the FLA as a Community of Practice (CoP, Wenger 1998) or discourse community (Swales 1990: 21-32) and interprets the practice of 'following' as a form of ambient affiliation. As such, this paper investigates the relationships between the FLA and its followership, seeing the relationship(s) between the FLA and its followers as co-constructive and negotiated dialectically; the identities of both are negotiated in relation to one another. In order to investigate this relationship, this study focussed on data collected about Twitter[13] accounts following the FLA's official Twitter account (@lads_alliance). Data were collected during October 2017 through Twitter's Application Programming Interface (API) which was accessed using the R[14] twitteR[15] package. Using twitteR, it was possible to generate a list of all @lads_alliance followers (n=15,114), which enabled the extraction of biographical data from all followers as well as tweets from any non-private followers (n=13,091).

Preliminary analyses of lexical terms in biographical descriptions suggests that followers predominantly identify with a football club (*fan, football, season ticket, Spurs, Arsenal*), but other identities also frequently feature, such as (gendered) relational identities (*dad, husband, married, wife*) and nationality (*British, English, Englishman*). Many profiles also contain reference to politics (*brexit, politics, ukip*), patriotism (*patriot*) and *religion*, specifically *Islam*.

---

[6] http://www.twitter.com/lads_alliance
[7] https://www.facebook.com/pg/FootballladsallianceFLA/about/?ref=page_internal
[8] http://footballladsalliance.co.uk/ABOUT%20US.html
[9] https://www.youtube.com/watch?v=ghF83NIufrw
[10] https://shysociety.co.uk/2017/10/07/anne-marie-waters-i-want-to-keep-the-britain-we-know/
[11] http://www.shariawatch.org.uk
[12] https://en.wikipedia.org/wiki/Pegida_UK
[13] https:// twitter.com
[14] https://cran.r-project.org
[15] https://cran.r-project.org/web/packages/twitteR/

An examination of hashtag usage in biographical descriptions again highlights themes of football club identification (*#lufc, #avfc, #millwall*), politics (*#brexit, #forbritain, #maga, #ukip, #bluehand, #voteleave*), nationality (*#england*), religion (*#athiest, #christian, #banislam*) and patriotism (*#patriot*).

Methods from Corpus Linguistics and Critical Discourse Analysis will be used to explore how these (collective) identities are signalled in FLA's Twitter followership and will go on to analyse linguistic patterns in tweets sent by followers to investigate correlations between the language used in tweets and users' biographies.

References

Swales, J. 1990. *Genre Analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Wenger, E. 1998. *Communities of Practice: learning, meaning and identity.* Cambridge: Cambridge University Press.

Zappavigna, M. 2017. "Twitter". In C. Hoffman and W. Bublitz (eds.) *Pragmatics of Social Media.* Berlin: Mouton De Gruyter. pp. 201-24.

A comparative study of intimate partner violence episodes in digital written media: News values through time

Sergio Maruenda-Bataller
Universitat de València
sergio.maruenda@uv.es

In the last thirty years, there have been important advances in the media coverage or discussion of intimate partner violence (IPV) (Aran Ramspott & Medina Bravo 2006; Vallejo-Rubinstein 2005). Lately, it is indisputable that IPV is one of the key issues not only in political, social and institutional discourses but also in the selection agenda of news producers. The recognition of this phenomenon has been largely due to the media, which have played a decisive role in transferring the issue from the private and personal to the public sphere, thus ensuring visibility and contributing to sensitizing citizenship (Berganza Conde 2003). However, some authors (e.g. Altés 1998; Alberdi & Matas 2002) have argued that this is not without a cost. Media are torn between two conflicting interests: on the one hand, to treat these grievous cases with the required ethics and, on the other, to attract a maximum audience, which is almost 'naturally' done through sensationalism. Journalists can create different pictures of domestic violence and "confirm and debunk the myths surrounding it by choosing certain topics, sources, facts, and words over others" (Bullock & Cubert 2002: 479).

Against this backdrop, this paper presents a corpus-assisted discourse study (CADS) of the news values used to discursively represent women victims of IPV in Spanish and UK dailies over the last decade, in an *ad-hoc* corpus of IPV news (2005-2016), in order to investigate whether news values have become more evaluative in time. Subsidiary to this, I will explore the way news values are exploited ideologically to construct discourse prosodies around women victims of IPV, violent episodes and perpetrators. The results gain insights into the social configuration and definition of women and their identities in contemporary written media on IPV through time, which are gathered in a pragmatic-discursive dictionary of key terms on gender violence.

For this aim, in this paper I apply Bednarek & Caple's (2014, 2017) *linguistic* approach to news values as discursive realisations of newsworthiness that "exist in and are constructed through discourse" (Bednarek & Caple 2014: 136). Following CADS (Baker & Levon 2015), linguistic occurrences are examined in order to gain insights into how news producers use news values to construct potential ideological discourses around IPV victims, paying attention to shared and different values cross-culturally, together with the most relevant discourse prosodies and ideological implications. Besides similarities in social and 'institutional' concerns, the qualitative analysis unveils significant contrasts. It shows that Negativity has more critical overtones in the Spanish newspapers, and reports on abusers are often constructed as more impersonal in the case of UK dailies. As for the depiction of extreme negative emotions, the higher number of occurrences, together with a wider plethora of word combinations construct Spanish reports as more ideological, if not sensationalist, thus exploiting readers' interest in crime and violence.

References

Baker, P, Gabrielatos, C. and McEnery, T. 2013. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press.* Cambridge: Cambridge University Press.

Baker, P. and Levon, E. 2015. "Picking the right cherries? A comparison of corpus-based and qualitative analyses of news articles about masculinity". *Discourse & Communication* 9(2): 221-236.

Bednarek, M. and Caple, H. 2014. "Why do news values matter? Towards a new methodological framework for analysing news discourse in Critical Discourse Analysis and beyond". *Discourse & Society* 25(2): 135-158.

Bednarek, M. and Caple, H. 2017. *The Discourse of News Values: How News Organizations Create Newsworthiness.* Oxford: OUP.

Bell, A. 1991. *The Language of News Media*. Oxford: Blackwell.

Bullock, C. and Cubert, J. 2002. "Coverage of domestic violence fatalities by newspapers in Washington State". *Journal of Interpersonal Violence* 17: 475-499.

Carter, C. and Steiner, L.. 2004. "Introduction to critical readings: Media and gender". In C. Carter and L. Steiner (eds.) *Critical Readings: Media and Gender*, 1-10.

Fagoaga, C. 1999. "*La Violencia en los Medios de Comunicación. Maltrato en la Pareja y Agresión Sexuada.*" Madrid (España): Dirección General de la Mujer de la Comunidad de Madrid.

Fowler, R. 1991. *Language in the News. Discourse and Ideology in the Press.* London & New York: Routledge.

Martin, J. and White, P. 2005. *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.

Meyers, M. 1997. *News Coverage of Violence Against Women: Engendering Blame*. Newbury Park: Sage Publications.

Montgomery, M. 2007. *The Discourse of Broadcast News: A Linguistic Approach.* London/New York: Routledge.

Partington, A. 2004. "Corpora and discourse: a most congruous beast". In A. Partington, J. Morley, and L. Haarman (eds.) *Corpora and Discourse*. Bern: Peter Lang: 11-20.

Vallejo Rubinstein, C. 2005. "*Representación de la Violencia contra las Mujeres en la Prensa Española (El País/El Mundo) desde una Perspectiva Crítica de Género: Un Análisis Crítico del Discurso Androcéntrico de los Medios.*" Unpublished MA thesis. Universitat Pompeu Fabra.

A look at male and female Late Modern scientific English from a Multidimensional perspective

Leida Maria Monaco
University of A Coruña
leidamaria.monaco@udc.es

Luis Puente-Castelo
University of A Coruña
luis.pcastelo@udc.es

The present study explores patterns of register variation in eighteenth- and nineteenth-century scientific texts by comparing male and female scientific discourse through the lens of Multidimensional Analysis, a now widely used methodology developed by Biber (1988) and successfully adopted in a large number of variation studies all over the world (Conrad, 1996; Lee, 2000; Gray, 2011; Grieve, 2014, among many others). Multidimensional studies use factor analysis, a multivariate statistical technique that permits to spot co-occurrence patterns and uncover underlying 'dimensions of variation' in a corpus, which serve as a reference to characterise the different texts and registers in that corpus. In this case, the texts analysed belong to three subcorpora of the *Coruña Corpus of English Scientifc Writing*, namely, the *Corpus of English Texts on Astronomy* (*CETA*), the *Corpus of English Philosophy Texts* (*CEPhiT*) and the *Corpus of English Life Sciences Texts* (*CELiST*) (Moskowich & Crespo, 2007; Lareo & Esteve-Ramos, 2007; Crespo & Moskowich, 2010; Moskowich & Crespo, 2012; Moskowich et al. 2016).

Bearing in mind that the Late Modern period was characterised by a rather explicit prejudice against the presence of women in science (Schteir, 1987, 2008; Trouille, 1997; Torralbo, 2010), there has been a number of studies on the *Coruña Corpus* which have focused on the language of female scientists (Lareo, 2011; Moskowich, 2013; Crespo, 2014; Puente-Castelo & Monaco, 2013, 2016; Moskowich & Monaco, 2014, 2016; Puente-Castelo, 2016). In this particular case, the aim is to characterise the discourse of female scientists –as opposed to that of male scientists– with respect to the four dimensions of variation found to date in the *Coruña Corpus* (Monaco, 2017): involved/persuasive vs. informational style, descriptive focus, discourse elaboration and narrativity. Dimension scores for male and female authors will be compared for each of the three scientific disciplines (astronomy, philosophy and life sciences), and for some of the genres (e.g. treatise, essay, lecture, etc.) present in the corpus and which include texts written by scientists of both sexes. Likewise, eighteenth- and nineteenth-century scores for both sexes will be contrasted to determine the extent of the presence of diachronic change in each case.

References

Biber, D. 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.

Conrad, S. 2001. Investigating academic texts with corpus-based techniques: An example from biology. *Linguistics and Education* 8, 299-326.

Crespo, B. 2014. Female Authorial Voice: Discursive Practices in Prefaces to Scientific Works. In M. Gotti & D. S. Giannoni (eds.) *Corpus Analysis for Descriptive and Pedagogic Purposes: English Specialised Discourse* (189- 202). Bern: Peter Lang.

Crespo, B. & Moskowich, I. 2010. CETA in the Context of the Coruña Corpus. *Literary and Linguistic Computing*, 25(2), 153-164.

Gray, B. 2011. *Exploring academic writing through corpus linguistics: When discipline tells only part of the story*. Unpublished PhD dissertation. Northern Arizona University.

Grieve, J. 2014. A Multi-Dimensional analysis of regional variation in American English. In T. Berber Sardinha & M. Veirano Pinto (eds.) *Multi-dimensional analysis, 25 years on: a tribute to Douglas Biber* (3-33). Amsterdam: John Benjamins.

Lareo, I. 2011. Colocacións con make, take e do + nome nun corpus do século XIX de textos ingleses científicos e literarios escritos por mulleres. *Cadernos de Fraseoloxía Galega* 13, 191-214.

Lareo, I. & Esteve-Ramos, M. J. 2007. 18th century scientific writing. A study of make complex predicates in the Coruña Corpus. *ICAME* 32, 69-96.

Lee, D. Y. W. 2000. *Modelling variation in spoken and written language: The Multi-Dimensional approach revisited*. Unpublished PhD dissertation. Lancaster University.

Monaco, L. M. 2017. *A Multidimensional analysis of Late Modern English scientific texts from the Coruña Corpus*. Unpublished PhD dissertation. University of A Coruña.

Moskowich, I. 2013. Eighteenth-century female authors: Women and science in the Coruña Corpus of

English Scientific Writing. *Australian Journal of Linguistics* 33(4), 467-487.

Moskowich, I. & Crespo, B. 2007. Presenting the Coruña Corpus: A collection of samples for the historical study of English scientific writing. In J. Pérez Guerra, D. González-Álvarez, J. L. Bueno-Alonso & E. Rama- Martínez (eds.) *Of Varying Language and Opposing Creed: New Insights into Late Modern English* (341-357). Bern: Peter Lang.

Moskowich, I. & Crespo, B. 2012. (eds) *Astronomy 'playne and simple': The writing of science between 1700 and 1900.* Amsterdam: John Benjamins.

Moskowich, I. & Monaco, L. M. 2014. Abstraction as a means of expressing reality: Women writing science in Late Modern English. In M. Gotti & D. S. Giannoni (eds.) *Corpus Analysis for Descriptive and Pedagogical Purposes* (203-224). Bern: Peter Lang.

Moskowich, I. & Monaco, L. M. 2016. Linking ideas in women's writing: Evidence from the *Coruña Corpus. Revista de Lingüística y Lenguas Aplicadas*, 11: 35-47.

Moskowich, I., Camiña, G., Lareo, I. & Crespo, B. 2016. *'The Conditioned and the Unconditioned': Late Modern English Texts on Philosophy.* Amsterdam: John Benjamins.

Puente-Castelo, 2016. Uses of Conditionals in the work of eighteenth and nineteenth century female scientists: applying quantitative methods to the study of female discourse in the Coruña Corpus. In I. Moskowich, B. Crespo & C. Núñez-Puente (eds), *Queering Women's and Gender Studies*. Newcastle-upon-Tyne: Cambridge Scholars Publishing.

Puente-Castelo, L. & Monaco, L. M. 2013. Conditionals and their functions in women's scientific writing. *Procedia: Social and Behavioral Sciences* 95, 160-169.

Puente-Castelo, L. & Monaco, L. M. 2016. The might of *might*: A mitigating strategy in eighteenth and nineteenth century female scientific discourse *Revista Canaria de Estudios Ingleses* 72 (143-168).

Schteir, A. 1987. Botany in the Breakfast Room: Women and Early Nineteenth-century British Plant Study. In P. Abir-Am and D. Outram (eds.) *Uneasy Careers and Intimate Lives: Women in Science* (1789-1979). New Brunswick, NJ: Rutgers University Press.

Schteir, A. 2008. "Elegant Recreations? Configuring Science-writing for Women". In B. Lightman (ed.) *Victorian Science in Context* (236–255). Chicago: University of Chicago Press.

Torralbo, J. 2010. Mujer y escritura en la Inglaterra de los siglos XVII y XVIII: Corte, autobiografía e reivindicación. *Anglogermanica Online* 7, 85-99.

Trouille, M. 1997. *Sexual politics in the Enlightenment: Women writers read Rousseau.* Albany, NY: Suny Press.

"Bad language" in film translation: a corpus-driven study

Effie Mouka
National and Kapodistrian University of Athens
efimouka@gmail.com

Ioannis E. Saridakis
National and Kapodistrian University of Athens & Lancaster University
i.saridakis@lancaster.ac.uk

The proposed paper reports on the findings of a corpus-driven study of the transfer of "bad language", "profane" or "taboo" words, in English to Greek film translation, in contrast to original film discourse. Our aim is thus to quantitatively investigate and then socially and register-wise interpret observed patterns (e.g., Mouka et al. 2015) in the translation of Bad Language Words (BLWs, McEnery 2006), in an as yet relatively under-researched area, that of film translation.

In line with McEnery's remark, in relation to English, that spontaneous spoken discourse is a "much more secure basis for the study of bad language" (2006: 28), and considering that in corpus-wise under-resourced languages such as Greek, the compilation and study of a systematic film corpus is perhaps the only reasonable resource for studying specific traits of oral discourse. Film discourse can in fact usefully represent the oral varieties of a language (Frochini 2012: 90).

In the field of AVT, previous research on the perception and transfer of "bad language" is, for the most part, qualitative (see, e.g., the work of Díaz-Cintas [2001] on the rendition of expressions with a sexual connotation, of Chen [2004] on the Hong Kong Chinese subtitling of English swearwords, of Scandura [2004] on (self-) censorship in subtitling, of Pujol [2006] on the transfer of fuck into Catalan, of Fong [2009] on the translation of vulgarisms and sexually-oriented language, of Greenall [2011] on the non-translation of swearing in subtitling as an example of loss of social implicature, and of Baines [2015] on the use of cues of register for triggering audience reactions to taboo language). This body of research seems to concur to the assumption that this type of language is often omitted or weakened in AVT (Fawcett 1997: 119; Díaz-Cintas & Ramael 2007: 195), accounting for and exemplified on grounds of technical (time and target-text length) contraints (e.g., Díaz-Cintas & Ramael 2007), or as a result of intersemiotic shift on the level of politeness, from an oral to a written mode (e.g., Gambier 2002, in Baines 2015: 437); one that implies downplaying or perhaps censoring "bad language", which in written discourse is presumably more striking than it is in an oral text: this is the case with film subtitles (see, e.g., Mayoral 1993: 57, in Díaz-Cintas 2001: 51).

By contrast, our work draws on McEnery's (2006: 1) postulate that "bad language" (and hence, its transfer into another socio-linguistic context, as is the case with AVT) should be investigated systematically, based on ample linguistic evidence, and be considered as a complex social phenomenon, so as to be able to explain "both the source of the undoubted power of bad language and the processes whereby inferences are drawn about speakers using it" (ibid.).

For this purpose, we have compiled: (a) a corpus of transcribed dialogues of five contemporary American films; (b) a corpus of their DVD (i.e. commercial) Greek subtitles; (c) a corpus of their Greek "fansubs" (a genre that is less or not at all subject to self-censorship and technical constraints); and (d) a reference corpus of transcribed dialogues of five contemporary Greek films with a marked use of "bad" language.

Corpora (a), (b) and (c) were aligned, with the aim to statistically investigate BLW-related findings. English BLWs, their corresponding renditions into Greek, and Greek BLWs from the original film corpus (d) have been annotated using McEnery's (2006: 27) categorisation schema. BWLs in the four annotated corpora have been quantitatively analysed (collocation- and concordance-wise), so as to assess:
(a) the tendency towards a more "standard", i.e. less "abusive" language in the rendition of BWLs in translated film language, i.e. in the two translation sub-corpora (b and c);
(b) the impact of (self-) censorship on film translation, by statistically comparing and interpreting the corresponding findings from corpora (b) and (c); and
(c) the "status" of translated film language in relation to original film language, by comparing the statistics and the classification of BLWs from corpora (b) and (c) with those of the original film corpus (d).

Not surprisingly, translated film language is found to be more "neutralised" when compared to "natural" film discourse. Thus, self-restriction (or self-censorship) is, indeed, a general pattern, which is to some extent eliminated in the case of fansubs, where subtitlers are not bound by the restraints of the text genre and of the AVT industry. However, research findings suggest that in film translation, the rendition of "bad language" is a considerably more intricate phenomenon. The lexico-grammatical and discourse profiling of Greek subtitles in contrast to original film language suggests that Pym's (2012: 108) "safety" postulate seems to be at play here, when

there are no "ready-made" correspondences. In other words, sociolinguistically, this profiling suggests a reluctance on the part of translators, one that significantly suppresses the source text's interpersonal metafunction (Munday 2012; Saridakis 2015: 206).

References

Baines, R. 2015. Subtitling taboo language: using the cues of register and genre to affect audience experience?. *Meta* 60.3, 431–453.

Chen, C. 2004. On the Hong Kong Chinese subtitling of English swearwords. *Meta* 49.1,135-147.

Díaz-Cintas, J. 2001. Sex, (sub)titles and videotapes. In L. Lorenzo García, and A.M. Pereira Rodríguez (eds). *Traducción subordinada II: el subtitulado (inglés-español/galego)*. Vigo: Universidade de Vigo, 47-65.

Díaz-Cintas, J., and A. Remael. 2007. *Audiovisual Translation: subtitling*. Manchester: St. Jerome.

Fawcett, P. 1997. *Translation and language. Linguistic theories explained*. Manchester: St. Jerome.

Fong, G. 2009. The two worlds of subtitling: the case of vulgarisms and sexually-oriented language. In K. Au, and G. Fong (eds). *Dubbing and subtitling in a world context*. Hong Kong: The Chinese University Press, 39-62.

Frochini, P. 2012. *Movie language revisited. Evidence from multi-dimensional analysis and corpora*. Bern: Peter Lang.

Gambier, Y. 2002. Les censures dans la traduction audiovisuelle. *TTR* 15.2, 203-221.

Greenall, A.K. 2011. The non-translation of swearing in subtitling: Loss of social implicature?. In A. Şerban, A. Matamala, and J-M. Lavour (eds). *Audio-visual translation in close-up. Practical and theoretical approaches*. Bern: Peter Lang, 45-60.

McEnery, T. 2006. *Swearing in English. Bad language, purity and power from 1586 to the present*. London: Routledge.

Mayoral, R. 1993. La traducción cinematográfica: el subtitulado. *Sendebar* 4, 45-68.

Mouka, E., I.E. Saridakis, and A. Fotopoulou. 2015. Racism goes to the movies: A corpus-based study of cross-linguistic racist discourse annotation and analysis. In C. Fantinuoli, and F. Zanettin (eds). *New directions in corpus-based translation studies*. Berlin: Language Science Press, 35-70.

Munday, J. 2012. *Evaluation in translation. Critical points of translator decision-making*. London: Routledge.

Pujol, D. 2006. The translation and dubbing of "fuck" into Catalan: the case of 'From Dusk till Dawn'. *Journal of Specialised Translation* 6, 121-133.

Pym, A. 2012. *On translator ethics. Principles for mediation between cultures*. Amsterdam: John Benjamins.

Saridakis, I.E. 2015. Probabilistic laws and risk aversion in translation: a case study in translation didactics. *Current Trends in Translation Teaching and Learning E (CTTL-E) 2015*, 196-245.

Scandura, G. 2004. Sex, lies and TV: censorship and subtitling. *Meta* 49.1, 126-134.

A corpus-based study of the rendition of contrastive markers in Chinese–English political interpreting

Jun Pan
Hong Kong Baptist University
janicepan@hkbu.edu.hk

Political interpreting, as part of political discourse (cf. Van Dijk, 2002; Schäffner and Bassnett, 2010), requires a high level of pragmatic competence (Arroyo, 2015). It is usually situated at the intersection of politics, media and language, which poses great difficulty as compared to other types of interpreting. Interpreters are often easily chosen "scapegoats", especially when frictions or problems occur (Buri, 2015). It would therefore be interesting to see how interpreters deal with the pragmatically high-risk linguistic indicators that lead to potentially unfavourable or unexpected contrasts. Will they choose to intensify the contrast posed by the linguistic code, or

to weaken the contrast? What would be context for them to decide which strategy to take? Not much research has been done to answer these questions.

It is under this background that the present study was conducted to investigate and examine the rendition of contrastive markers (CMs) in Chinese–English political interpreting. The study employs the definition of Fraser (1996, p. 168, 187), which specifies CMs as "the linguistically encoded clues" that signal "a denial or a contrast of some proposition associated with the preceding discourse". Two representative CMs in Chinese are 不過 (suggesting concession, a "lighter" version of contrast) and 但是 (indicating denial, a "stronger" version of contrast) in Chinese, equivalent in meaning to *however* and *but* respectively in English, were chosen for analysis. Two parallel corpora were constructed for the purpose of the study: one composed of the annual Reports on the Work of the Government of the People's Republic of China (PRC) and their interpretations in English, and the other including Policy Addresses delivered by the Chief Executives in the Hong Kong SAR on an annual basis and their English interpretations. The speeches represent two varieties of the Chinese language (i.e., Putonghua and Cantonese), and their renditions two types of interpreting (i.e., consecutive and simultaneous interpreting). The corpora were aligned at the paragraph level. A semi-automatic process was employed to identify the use of 不過 and 但是 in the Chinese part of the corpora, after which, their renditions in English were manually extracted.

Findings of the study suggest that apart from being rendered into their total equivalence, i.e., *however* and *but*, both 不過 and 但是 were interpreted into different CMs in English. The renditions of the two CMs could be roughly categorized into four cases: 1) intensification (a CM is rendered into a "stronger" version, e.g., interpreting 不過 into *but*), 2) total equivalence; 3) mitigation (a CM is rendered into a "lighter" version, e.g.., interpreting 但是 into *however* or *nevertheless*); and 4) omission (there is no equivalence of the CM in the interpreted version). The study shows that total equivalence seems to be the most common strategy employed by interpreters, which was then followed by mitigation or omission. These strategies were then examined in the contexts of simultaneous and consecutive interpreting, and of Putonghua and Cantonese source texts.

The study shows the preferred strategies employed Chinese–English political interpreters in the rendition of CMs, a set of pragmatically high-risk pitfalls, in different contexts. Its findings will provide insights into the training of political interpreters.

References

Arroyo, J. L. B. 2015. "Pragmatics of political discourse". In C. A. Chapelle (ed.), *The encyclopedia of applied linguistics*. Hoboken: John Wiley & Sons.

Buri, M. R. 2015. "Interpreting in diplomatic settings". aiic.net. Retrieved from http://aiic.net/p/7349.

Fraser, B. (1996). "Pragmatic markers". *Pragmatics*, *6*, 167–190.

Schäffner, C. and Bassnett, S. 2010. "Politics, media and translation: Exploring synergies". In C. Schäffner & S. Bassnett (eds.) *Political discourse, media and translation*. Newcastle: Cambridge Scholars Publishing.

Van Dijk, T. 2002. "Political discourse and ideology". In C. U. Lorda & M. Ribas (eds.) *Anàlisi del discurs politic*. Barcelona: Universitat Pompeu Fabra, pp. 15-34.

National face and facework. A corpus-assisted case study of Chinese and US press briefings

Alan Partington
Bologna University
alanscott.partington@unibo.it

Maria Marakhovskaiia
Bologna University
maria.marakhovskaiia@studio.unibo.it

The Goffman / Brown & Levinson socio-pragmatic theory of face was first devised through observing the interaction of individuals. Later research has looked at the phenomenon of group-face (e.g. Spencer-Oatey 2007). Here we examine whether face theory can be applied to communications made by state actors to the outside world, in other words, whether facework theories could also be applied to *national* face.

Chinese government statements are a promising discourse type for the study of national facework mainly because of the intricate relations China has with its geographical neighbours, but also because of the supposed special importance accorded to face in Chinese culture (Chen & Hwang 2016). Moreover, the lion's share of corpus linguistic studies on media briefings has been conducted on Western English-language data.

According to Brown & Levinson (1987), face consists of two complementary facets, negative and positive face. Negative face is the desire for freedom from imposition and freedom of action. It is 'respect my space' face and thus negative facework is respecting the space of others and – crucially here – trying to ensure others respect your face and freedom of action. Positive face, instead, involves the desire to be respected, approved of and admired, a notion similar to *ethos*, Aristotle's first part of rhetoric, 'how the speaker/ writer projects their personality and stance towards the audience / readership' (Taylor 2010: 222).

Partington has also indicated how, in institutional settings at least, positive politeness can be of two different kinds, namely, *competence* face and *affective* face (2006: 97–98 et passim). One's competence face is one's image as well-informed, expert, in control and authoritative. One's affective face is one's image as likeable, good-humoured, normal, 'one of us'. However, a major problem for institutional speakers is that the two kinds of face are not always compatible; it is not always possible to project an image of authority and expertise at the same time as one of a normal, easy-going person and speakers need to know when to prioritise one over the other. All these types of face were objects of study here.

We compiled a corpus of transcriptions in English translation of all the press conferences held by the Ministry of Chinese Foreign Affairs in 2016 (236 conferences, 270,000 words).[16] The Chinese language versions are reported here alongside the English.[17] This was subjected to a mixture of quantitative and qualitative analysis, as well as comparative analysis with US White House press briefings.

As regards the quantitative analysis, the most productive methods turned out to be cluster, (n-gram), frequency listing. Lists of 1- to 5-grams were prepared, followed by concordancing of items which either appeared significant from the cluster listing or which were fairly obvious candidates for examination, such as the names of neighbouring countries or of powerful nations

---

[16] The English language versions are vailable at:
www.fmprc.gov.cn/mfa_eng/xwfw_665399/s2510_665401/2511_665403/
[17] The first author is proficient in Chinese.

interacting with China. Collocation lists deriving from these concordances were also examined. WordSmith Tools (Scott 2008) was used for all these procedures. The comparative corpus of White House press briefings WH-2011, also contains all the briefings of the respective administration over the course of a year, from December 2010 to the end of November 2011, a total 239, comprising 1,300,000 words.

As regards the contrastive research, the most striking finding was how transactional, formal and formulaic the CFA briefings are relative to those conducted at the White House or US State Department (Jiang 2006), with little interpersonal content. The questions generally take the form of a statement on some foreign policy issue, followed by a formally interrogative locution, such as 'what is your / China's comment on that?' 请问中方对此有何评论 (390 occurrences), 'Do you have any comment (on this/that)' 你对此有何评论? (63 occurrences), 'What is your / China's response (to that/this)?' 中方对此有何回应?(148). Moreover there are no names used, the podium much prefers to use *we*我们(1,891 occurrences) than *I*我 (834), stressing their role as representative rather than individual human interlocutor (Marakhovskaiia 2016). There are no follow-up questions, so frequent in the White House briefings and certainly none of the banter and other kinds of affective facework which permeate the latter (Partington 2003, 2006). They are openly geared to doing the business of protecting and enhancing China's national face in the world. There are, then, stark differences between the rhetorical strategies pursued by the podiums in US and Chinese briefings as regards competence and affective face.

There are, however, also similarities between Chinese and US briefings. Opportunities are seized by both sets of podiums to project the positive face of the home administration and often the values expressed are similar, namely, being peace-loving, responsible and cooperative. Both bolster their own positive face by claiming to represent international opinion and good sense. Both prize their own negative face and their freedom of action very highly. Both are ready to impose heavy imposition on the negative face of others, to demand they change behaviour, and both issue threats of potential bad consequences. The commonplace that face is particularly important in Chinese culture was therefore not proven, at a national level at least. It seems more likely that, although some of the details may alter, facework is similarly fundamental, unavoidable and necessary in both, and perhaps all, cultures.

References

Brown, Penelope & Stephen Levinson. 1987. *Politeness: Some universals in language usage*. Cambridge: Cambridge University press.

Chen, Ron & Kwang-Kuo Hwang. 2016. Nation, Face, and Identity: An Initial Investigation of National Face in East Asia. Frontiers in Psychology, 7:1557, 1-11.

Jiang, Xangying. 2006. Cross-cultural pragmatic differences in US and Chinese press conferences: the case of the North Korea nuclear crisis. *Discourse & Society* 17(2) 237–257.

Marakhovskaiia, Maria. 2016. China's Foreign Policy as Represented in Domestic and Western Sources. M.A. thesis, School of Political Science, Bologna University.

Partington, Alan. 2003. *The Linguistics of Political Argument: The Spin-doctor and the Wolf-pack at the White House*. London: Routledge.

Partington, Alan. 2006. *The Linguistics of Laughter. Laughter-talk at the White House*. London: Routledge.

Scott, Michael. 2008. *WordSmith Tools version 5*, Liverpool: Lexical Analysis Software.

Spencer-Oatey, Helen. 2007. Theories of Identity and the Analysis of Face. *Journal of Pragmatics* 39(4). 639–656.

Taylor, Charlotte. 2010. Science in the news: a diachronic perspective. Corpora 5(2) 221-250.

Blaming mothers for feckless fathers? Analysing reader comments posted in direct response to UK online newspaper articles

Laura L Paterson
The Open University, UK
laura.paterson@open.ac.uk

The comments feature often activated on online newspaper articles constitutes an important site of public debate. Allowing readers to respond directly to a stimulus text, the comments function encourages readers to discuss the topic of an article amongst themselves. The analysis of these reader comments using corpus-based discourse analysis can shed light on the prevailing ideologies members of the public associate with a particular issue. In this case, the focus is on the discussion and evaluation of mothers (and women more generally) in comments responding to articles about 'feckless' fatherhood in the UK. Paterson (in prep) showed that mothers – in particular single mothers – were blamed for paternal fecklessness by public commenters responding to a newspaper article about unemployed fathers in 2010. To establish whether this negative evaluation of mothers is a core component of a feckless fatherhood discourse, or represented an isolated case, this paper expands the study of publicly-authored comments to those posted under relevant articles from 2013-2018.

To create a corpus of public comments, Nexis UK was used to obtain a list of potential online articles using the following search criteria: <Feckless father OR Feckless fathers>, In the Headline, Previous 5 years (from 19/01/2018), UK publications, Duplicates removed. The search results were manually checked to filter out only those articles which had generated 100 or more public comments. There were 17 sets of comments which met this threshold (9731 comments). Although all UK national newspapers were included in the original Nexis search, most of the articles generating 100+ comments appeared on the *Mail Online* website, suggesting that this particular publication was a primary site for debating feckless fatherhood and/or the newspaper most likely to use this term. (Relevant articles from the *Guardian* and *Express* also had comments enabled, but tended not to reach the 100+ threshold.)

Articles using the term 'feckless father' either referred to generic paternal fecklessness or focused on specific examples of real men. The comments were thus split into two subcorpora (comments on articles concerning generalised feckless fatherhood OR individual feckless fathers). The analysis of these subcorpora addresses the following research questions:

1. What discourses are drawn upon by members of the public responding to online newspaper articles about feckless fatherhood?
2. Do these discourses differ based on the grammatical definiteness of the feckless father(s) in question?
3. How are mothers evaluated in the public comments?
4. What are the wider social implications of these evaluations of motherhood in the UK?

It is hypothesised that articles about generalised feckless fatherhood will prompt the most negative attitudes to mothers, as there is no specific target in the body of the original newspaper article. By contrast, articles about particular feckless fathers are expected to generate fewer negative comments about women, as the agency of the father is clear. The overall aim is to investigate how these two types of articles influence debates about feckless fatherhood and

whether there is a wider impact on attitudes to mothers and women in general.

References

Paterson, L.L. in prep. "Electronic supplement analysis of multiple texts: Exploring discourses of UK poverty in Below the Line Comments."

'No mencion made of us': Discursive Strategies and Rebellion in the Wars of the Roses

Tamara Peeters
Birmingham City University
tamara.peeters@mail.bcu.ac.uk

This paper focuses on a collection of seven texts produced by the Earl of Warwick, the Duke of Clarence and King Edward IV in 1469-1470. These texts are all part of a public campaign to reinstate Henry VI on the English throne, over the current king, Edward IV. The focus of the analysis is the discursive strategy of naming, and specifically the way the authors referred to themselves and their direct opponents, as this was a strategy that was used to self-identify their position in the social hierarchy, especially in relation to their king. Corpus analytical software was deployed to generate quantitative data relating to the texts, which was then subjected to a qualitative analysis using critical discourse analysis. Whereas previous research into this period treated these texts as the primary sources for historical events, which they are, the discourse has rarely been the focus of any previous studies, and even then these studies focus on discourse in literary texts (Leitch, 2015). By focusing on the discourses of power in these texts, it allows for innovative new insights into the power relations of the late fifteenth century and how they were established and maintained.

That language could be used as a means of manipulation during a conflict was not a strange concept, even in late fifteenth century England. While the Wars of the Roses have been intensively studied for years, no in-depth linguistic analysis of the interplay between language and power has taken place. Studies (e.g. Ross, 1981) have been carried out into propaganda during the Wars of the Roses, but they tend to focus on the historical perspective over a historical linguistic perspective.

The analysis of this collection of texts found that Warwick and Clarence rarely use positive ways to refer to themselves in relation to their allegiance to King Edward IV. In fact, Edward IV is mostly absent from the narrative overall. The moments positive terminology is used for self-reference, e.g. by using LIEGEMAN, SUBJECT or SERVANT, this is done to define the Earl of Warwick's role in relation to Margaret of Anjou and Henry VI, but not the current king Edward. The show of deference for the former king, but not the current king is very unusual. As all of these texts were written for the public sphere, the fact that two people who were always closest to King Edward IV diverge from the standards of the time did not go unnoticed by the monarch. In fact, Edward IV commented on this in a letter addressed to his brother, the Duke of Clarence, which indicates to us that the use of the aforementioned terminology was expected and part of the standard discourse at the time when addressing issues of state.

The analysis shows that Warwick defers to Henry VI in a way that he had not deferred to

Edward IV in any of these documents, which further affirms the enmity between the two in the late 1460s. Especially the fact that Edward IV picked up on it, shows that the lack of any formal deferrals to their monarch when speaking about matters relating to power is highly unusual for the time, and thus extra powerful to use as a discursive tool when trying to discredit those in power.

In addition, this paper exemplifies how CADS can offer a different perspective to history, thereby adding to our knowledge of the period, as well as help identify discursive strategies used hundreds of years ago that may still be in use today. Although CADS is a powerful tool to get a new perspective on history, it is rarely used to look at early versions of the (English) language. This study shows how valuable this methodology can be, and how the interdisciplinary approach of combining historical research with critical discourse analysis can contribute further to our knowledge of political discourse, history and how power relations were established and maintained historically, which allows for comparisons to contemporary politics and political discourse.

References

Leitch, M.G., 2015. *Romancing Treason* 1st ed., Oxford: Oxford University Press.

Ross, C., 1981. Rumour, Propaganda and Popular Opinion During The Wars of The Roses. In R. A. Griffiths, ed. *Patronage, The Crown and the Provinces in Later Medieval England*. Gloucester: Sutton Publishing Ltd, pp. 15–32.

Immigration through the lens of the Spanish and British judiciary. Bringing systemic linguistics into the equation

María José Marín Pérez
Universidad de Murcia
mariajose.marin1@um.es

According to the Eurostat[18], the statistical office of the European Union, there were 34.3 million people born outside the EU living in European countries on 1 January 2015. The largest numbers could be found in Germany (7.5 million people), United Kingdom (5.4 million), Italy (5 million), Spain (4.5 million) and France (4.4 million), 76% of the total amount of non-EU nationals in the whole of the EU.

Regrettably, non-EU immigration has recently become relevant owing to the humanitarian crisis of Syrian refugees, fleeing from the horrors of civil war. However, the response of the European authorities to this and other crises has been more than questionable, particularly in Spain, where less than 2,000 Syrian refugees have been granted asylum to date and the African borders are being invigilated applying strongly restrictive policies, which often go against the basic principles of human rights.

But how do legal systems react to this or other migration flows into the EU? Could this be observed and quantified in legal text? How may these observations differ depending on the legal system these texts stem from? Could systemic linguistics become part of the equation in such analysis?

---

[18] http://ec.europa.eu/eurostat

This research presents a corpus-based study of two corpora of judicial decisions, issued by British and Spanish courts, which deal with the phenomenon of immigration. Its major aim is to try and find linguistic evidence of the way in which propositional information is expressed in this legal genre through the application of the appraisal framework as defined by White (1999), Martin (2003), Eggins and Slade (1997), Rothery and Stenglin (2000) and Kaltenbacher (2006), within the systemic linguistics theoretical framework.

The texts in each corpus were gathered using two databases: the "Bailii"[19] (the British and Irish legal Information Institute), which offers access to nearly 300,000 British and Irish primary legal documents, and the "Cendoj"[20] search engine, a free database comprising Spanish judicial decisions made by all the courts within the Spanish legal system. They were obtained by introducing the keywords immigrant and refugee in the English corpus and inmigrante and refugiado in the Spanish one without applying any filters to the search, all the decisions containing these terms were selected and downloaded.

The reasons to apply the systemic linguistics appraisal framework to the study of these text collections are basically related to the capacity of such framework to cater for the expression of engagement and stance on the part of the author, and thus reveal certain attitudes to the content of the text.

Given the differing nature of the British and Spanish legal systems, the former emanating fundamentally from case law, the latter stemming from statutes and codes like most civil law countries, our initial hypothesis was that the results would show a greater inclination on the part of the British texts to explicitly reflect the authors'/speakers' engagement/stance towards the topic more overtly (frequency counts and the identification of collocate patterns could account for this fact). Conversely, the Spanish texts would tend to be more aseptic and neutral precisely because of their lesser dependence on case law, which necessarily involves referring to other similar cases, where the human and personal component may add to the parties' arguments and the judge's reasoning as well as to the final decision itself

In order to process both corpora in search for evidence to confirm this hypothesis both Scott's (2008) Wordsmith Tools and Brezina et al.'s (2015) Graphcoll software were used. The initial findings appear to confirm our hypothesis. Our proposal will include all the results and a full discussion of the major findings aiming at throwing some light on the overall vision of immigration that the Spanish and British judiciary convey in their decisions.

References

Alonso, A., Millon, C. & Williams, G. (2011). **"**Collocational networks and their application to an E-Advanced Learner's Dictionary of Verbs in Science (DicSci)". Kosem, I., Kosem, K. (eds.) In *Proceedings of E-lex 2011.* Ljubliana: Trojina, Institute for Applied Slovene Studies.

Bowker, L. & Pearson, J. (2002) *Working with Specialized Language: A Practical Guide to Using Corpora.* London: Routledge.

Baker, P., Gabrielatos, C. & McEnery, T. (2013). "Sketching Muslims: A corpus-driven analysis of representation around the word "Muslim" in the British press, 1998-2009. *Applied Linguistics*, 34(3), 255–278.

Eggins, S. & Slade, D. (1997). *Analysing Casual Conversation.* London: Casell.

---

[19] http://www.bailii.org/
[20] http://www.poderjudicial.es/search/indexAN.jsp

Kaltenbacher, M. "Culture-Related Linguistic Differences in Tourist Websites: the Emotive and the Factual. A Corpus Analysis within the Framework of Appraisal". In G.Thompson and S. Hunston (eds.). *System and Corpus. Exploring Connections.* London: Equinox.

McEnery, T. & Hardie, A. (2011). *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.

McEnery. T. & Wilson, A. (1996). *Corpus Linguistics.* Endinburgh: Edinburgh University Press.

Martin, J.R. (2003). "Beyond Exchange: APPRAISAL Systems in English". In S. Hunston and and G. Thompson (eds.). *Evaluation in text. Authorial Stance and the Construction of Discourse.* Oxford: Oxford University Press.

Rothery, J. & Stenglin, M. (2000). "Interpreting Literature: the Role of APPRAISAL". In L. Unsworth (ed.). *Researching Language in Schools and Communities: Functional Linguistics Perspectives.* London: Casell.

Sinclair, J. (2005). "Corpus and Text-Basic Principles". M. Wynne (ed.). *Developing Linguistic Corpora: a Guide to Good Practice.* Oxford: Oxbow books.

White, P.R.R. (1999). *The Language of Attitude, Arguability and Interpersonal Positioning.* The appraisal website: Homepage. http://www.grammatics.com/appraisal/

From keywords to discourse – towards a systematic approach

Joachim Peters
University of Erlangen-Nuremberg
joachim.peters@fau.de

Natalie Dykes
University of Erlangen-Nuremberg
natalie.mary.dykes@fau.de

Keywords are widely used as the first entry point into corpus-based discourse analysis (Baker et al. 2013). However, it is unclear how they should be calculated and how to categorise them in a systematic and linguistically motivated way (Baker 2004). Our approach combines three different keyness measures and suggests a categorisation procedure that is based on linguistic rather than sociological categories and therefore topic-independent.

Our analysis is based on a 1.3M token corpus of German press texts about multi-resistant pathogens. A thematically similar, much smaller corpus has been qualitatively analysed in a previous study (Peters 2017). Drawing on the results, we develop a linguistic categorisation scheme to classify discursive entitities such as actors, metaphors and argumentation topics. Our corpus study attempts to evaluate these qualitatively determined categories through quantitative evidence.

While log likelihood is known for highlighting high-frequency keywords (Hardie 2014), effect-size measures tend to be biased towards low frequency (ibid.). We thus combine three measures, extracting words with various frequencies (log likelihood, log ratio and a conservative estimate for the log ratio coefficient using the lower bound of a 99% confidence interval with Bonferroni correction) and pool the keywords for annotation. We apply two German national broadsheet newspapers as reference corpora:

- years 2011–2014 of *Süddeutsche Zeitung* (SZ), a left-leaning daily newspaper (290M tokens)
- years 2011–2014 of *Frankfurter Allgemeine Zeitung* (FAZ), a right-leaning daily newspaper (150M tokens)

Incorporating concepts from established models of qualitative discourse analysis (cf. DIMEAN by Spitzmüller & Warnke 2011 for an extensive and well-known example), we develop a processual model for systematic keyword exploration:

1	Calculation of keywords through various measures	↓
2	Group keywords using linguistic categories	↓
3	Explore the various levels of granularity in said categories yielded by the different keyness	↓
4	Examine the categories more closely through a sample of concordances	

➔ Goal: identification and manual deepening of discourse patterns in large amounts of specific text

Our gold standard is a scheme of 33 categories identified in a previous qualitative discourse analysis (Peters 2017). It comprises various topics, metaphorical source domains, actors and evaluative lexis. All 455 keywords are independently annotated by the two of us, resulting in an agreement of 82.2%. Following a comparison of annotations, we conclude that the disagreements are to a large extent systematically motivated:

- one annotator identified a topos not present in the gold standard scheme (these words were marked as false positives by the other annotator)

- some of our categories exhibit a significant amount of content overlap (topic expressing a problem vs. its solution)

The measures yield around the same number of true/false positives. Thus, the quality of keywords is roughly the same, but the words highlight different levels of discourse

All keyness measures yield a number of words that fit most actors and a number of topics. Metaphorical source domains are underrepresented in all of the keyword lists. This can partly be compensated through examination of KWIC: war metaphors, only represented by the word *bekämpfen* (to fight) frequently form clusters, so that concordances yield words like *Waffe* (weapon), *Angreifer* (attacker) and *Feind* (fiend).

Concerning differences between measures, both variants of log ratio generate much more specific keywords than log likelihood, thus providing a more detailed glimpse of the discourse it points to (cf. Pojanapunya & Watson Todd). For instance, log likelihood yields words like *doctor* and *nursing staff* as members of the *medical staff* actor category, while log ratio results include *hygiene specialist* and *hygiene inspector* (see table 1).

While log ratio shows more specific and thereby more concrete "entry points" into our data, log likelihood keywords are still valuable. They have a tendency to reflect the aboutness of more general topics which are also present in the reference corpus; in our case *medical press discourse*, and thus allow the researchers to examine concepts from a higher-level discourse in the context of the specific topic in question.

Discourse category	KW *Log Ratio*	KW *LR with confidence interval*	KW *Log likelihood*
actor: medical staff	deaconess, hygiene specialist, hygiene inspector	deaconess, hygiene specialist, hygiene inspector, hospital staff	doctor, family doctor, medic, nursing staff
metaphor: space	entry port, germ attack, pharyngeal	germ attack	spread, colonisation
metaphor: technical	biofilm, source of infection (lit. infection stove), infection chain, cell wall	biofilm, blood stream, source of infection (lit. infection stove), infection chain, cell wall	immune system

Table 1: Translated equivalents of keywords from different discourse categories, separated by keyness measure

References

Baker, P. (2004). Querying keywords - questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, *32*(4), 346–359.

Baker, P., Gabrielatos, C., & McEnery, T. (2013). *Discourse analysis and media attitudes: The Representation of Islam in the British Press*. Cambridge: Cambridge University Press.

Hardie, Andrew (2014). A single statistical technique for keywords, lockwords, and collocations. Internal CASS working paper no. 1, unpublished.

Peters, J. (2017). *Den Feind beschreiben. Multiresistente Erreger im deutschen Pressediskurs. Eine diskurslinguistische Untersuchung der Jahre 1994–2015* (master thesis). Friedrich-Alexander-University Erlangen-Nuremberg, unpublished.

Pojanapunya, P., & Watson Todd, R. (2016). Log-likelihood and odds ratio: Keyness statistics for different purposes of keyword analysis. *Corpus Linguistics and Linguistic Theory*. Retrieved from https://www.degruyter.com/downloadpdf/j/cllt.ahead-of-print/cllt-2015-0030/cllt-2015-0030.xml

Spitzmüller, J./Warnke, I. H. (2011): Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse. Berlin/New York.

The discursive construction of trolling and trolls on British political blogs based on the motives attributed to the trolls

Márton Petykó
Lancaster University
m.petyko@lancaster.ac.uk

This corpus-assisted study investigates a prominent social phenomenon of computer-mediated communication: trolling (Hardaker 2013). It aims to identify the linguistically marked motives that participants attribute to those they call trolls in the comment threads of British political blogs. The paper also focuses on how the motives attributed to trolls contribute to the discursive construction of trolling and trolls in the comments. Another goal of the paper is to examine whether the chiefly emotional motives, such as boredom (Baker 2001), loneliness (Fichman & Sanfilippo 2015), or hate towards other participants (Herring et al. 2002), ascribed to trolls in the academic literature correspond with those motives the participants attribute to the alleged

trolls.

The corpus consists of 1,712 comment threads. These threads were published on 27 British political blogs in 2015 and they include 740,841 comments. The size of the corpus is around 32.2 million tokens. 84% of the threads have been collected from seven blogs, *Conservative Home, Guardian Politics Blog*, *Guido Fawkes*, *LabourList*, *Left Foot Forward*, *Political Betting*, and *Wings over Scotland*, which therefore are the key British political blogs for this study. Every collected thread contains at least one comment in which a participant calls at least one other participant a troll or describes at least one other comment as an act of trolling, using a word form of the lexeme TROLL, such as *trolls, trolling*, or *troller*. I refer to these as 'troll comments'. In total, 6,129 troll comments have been identified in the corpus.

Data analysis involved a corpus-assisted qualitative-interpretative analysis of the collected troll comments: (1) Using the concordance lines of the search term **troll** in the software package *AntConc* (Anthony 2016), I selected and annotated those 2,459 troll comments in which the participants also discussed the possible reasons why the alleged trolls were trolling. I refer to these as 'troll motive comments'. (2) I identified the linguistically marked motives that participants attributed to those they called trolls and created a taxonomy from them. (3) I described how trolling and the trolls are constructed in the troll motive comments, depending on the motives attributed to the trolls. (4) To determine how often the participants explicitly attribute the identified motives to the alleged trolls, I used the motives as descriptive categories and provided each troll motive comment with motive-related annotations. (5) To make this discursive-pragmatic annotation process more transparent and consistent, I studied the n-grams and collocates of the search term **troll** and the positive keywords in the troll comments against the other comments in the threads using *AntConc*. The aim of this step was to identify those words and multiword expressions that mark a motive for trolling on their own. (6) Finally, I summarised the quantitative results of the annotation.

The main conclusions of this study are as follows:
 (1) Participants attribute five motives to the alleged trolls in the analysed troll motive comments. These are interrelated emotional/mental health-related/social reasons, financial gain, political beliefs, being employed by a political body, and an unspecified political affiliation.
 (2) Depending on these motives, trolling and the trolls are constructed in various ways. Trolling is constructed as (a) an emotionally, psychologically, and socially motivated, individual pastime activity, (b) a financially motivated individual activity, (c) an ideologically driven individual behaviour, or (d) a politically and financially motivated, centrally organised and comprehensively planned collective activity.
 (3) Similarly, trolls are portrayed as (a) socially deprived, unsuccessful, and traumatised individuals, who suffer from various mental health issues and/or emotional problems, (b) unskilled, low-paid, and financially deprived employees, (c) intellectually inferior, delusional, ignorant, intolerant, and extremist political fanatics, or (d) incompetent, exploited, and powerless employees of a political body.
 (4) Although the relevant academic literature regards trolling as a chiefly emotionally motivated behaviour, in the examined corpus, an unspecified political affiliation, being employed by a political body, and political beliefs are more frequently mentioned motives for trolling than emotional reasons.
 (5) Whilst trolling can be constructed in different ways in the analysed troll motive comments, a common trait of these comments is that the alleged trolls are portrayed in a strongly negative

manner. Thus, when calling others trolls, participants attribute motives to the trolls not only to explain their behaviour but also to belittle and discredit them.

References

Anthony, L. 2016. *AntConc* (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/

Baker, P. 2001. "Moral Panic and Alternative Identity Construction in Usenet". *Journal of Computer-mediated Communication*, 7 (1).

Fichman, P. and M. R. Sanfilippo. 2015. "The Bad Boys and Girls of Cyberspace: How Gender and Context Impact Perception of and Reaction to Trolling". *Social Science Computer Review* 33 (2): 163–180.

Hardaker, C. 2013. ""Uh….not to be nitpicky,,,,but...the past tense of drag is dragged, not drug.". An Overview of Trolling Strategies". *Journal of Language Aggression and Conflict*, 1 (1): 58–86.

Herring, S. C. et al. 2002. "Searching for Safety Online: Managing „Trolling" in a Feminist Forum". *The Information Society* 18 (5): 371–384.

Religion, Immigration and Integration: Self-representation amongst immigrant communities with religious identity

Chakravarthi Ram-Prasad
Lancaster University
c.ram-prasad@lancaster.ac.uk

Carmen Dayrell
Lancaster University
c.dayrell@lancaster.ac.uk

Gwen Griffith-Dickson
The Lokahi Foundation
ggd@lokahi.org.uk

This paper will examine self-representation amongst various immigrant communities with religious identity, living in the UK and economically integrated. To date, studies looking across religious traditions are rare. Most of the literature on immigration has focused either on race (Small 2006), or on the 'native' British responses to immigration (Winder 2005), or more recently, on Islam only (Strabac and Listhaug 2008, Moghissi and Gorashi 2010). This study differs from previous research in various ways: i) looks across different groups coded by both race/ethnicity and religion; ii) gives access to immigrants' views on their experiences and life in the context of immigration and religious values/identity; and iii) brings out the socio-economic success these individuals and communities have had in moving to and establishing themselves in Britain, and the role their religion has played in this transition.

The data analysed in this paper comprises 73 interviews (357,788 words) with immigrant communities. These interviewees were conducted in 2005 as part of a project commissioned by the Home Office. The original research was designed to follow conventional social science methods with respect to data collection and analysis. Set question areas were asked during the interviews for consistency across all groups but the interviewers' style was relaxed, informal, and conversational to ensure good rapport for methodological as well as ethical reasons. The interviewees were selected as 'success stories' in social or economic terms, and fall into four major groups, namely: (i) Christians; (ii) Hindus; (iii) Muslims; and (vi) a set of high achievers, either high net-worth individuals or outstanding in their fields for artistic, cultural or other standards. They are from various backgrounds and with various faiths, referred to internally (only) and casually as a 'SuperGroup'. The data collection from the SuperGroup did not differ from other groups... As 'success stories', they may feel most integrated, and thus more likely to espouse the language of British values. Alternatively, they may feel less pressure to comply with British values, practices and beliefs, or may even question those which are not consistent with their own. Christians were mainly African-Caribbean; Hindus and Muslims were in their vast majority from India, but there were some from surrounding countries (Pakistan,

Bangladesh, Burma, and Iraq). Within the SuperGroup, the majority was Muslim and originally from the Indian subcontinent; the exception being a Sikh from India and a Caribbean-born Hindu.

For the analysis, we first generated keywords and the key semantic domains (Rayson 2008) to obtain a broad overview of the interviewees' discourse, using WMatrix (Rayson 2009). This initial quantitative analysis provided the starting point of the analysis by identifying the most salient words that merited closer investigation. These selected words were examined through close reading of concordance lines, going as far back or forward in the text as it deemed necessary. In many cases, it was essential to consider the interviewers' question and/or comment to interpret the discourse.

The analysis provides evidence to support the claim that these interviewees feel well integrated into the British society irrespective of their religious beliefs, ethnicity or socio-economic status. While showing strong links with their family roots and background, they see themselves as British, irrespective of whether they were born in Britain. We also found that religion plays a key role in the interviewees' lives, shaping and influencing their behaviour, practices and relationships. Many do not see religion in isolation but intertwined with one's identity, culture, and family values. This perhaps explains the strong sense of community that prevails across all groups. Interviewees frequently mentioned how they engage with and contribute to those who share their religious beliefs, culture, and traditions. Most interviewees also expressed close connection with their families, making reference to their family histories, traditions, values, and religious beliefs and practices. At the same time, the interviewees seemed aware of their dual position in bridging and negotiating between cultures. They also mentioned that they interact with people from other religions, cultures and backgrounds, and the British community specifically. Many stressed that it is also essential to integrate and contribute to the British community at large.

References

Moghissi, H. and H. Gorashi (2010). *Muslim diaspora in the West: negotiating gender, home and belonging*. Farnham: Asghate.

Rayson, P. (2008). 'From key words to key semantic domains'. *International Journal of Corpus Linguistics* 13(4): 519-549.

Rayson, P. (2009). *Wmatrix: a web-based corpus processing environment*. Lancaster, Computing Department, Lancaster University.

Small, S. (2006). 'Race, Immigration and Politics in Britain: Changing Policy Agendas and Conceptual Paradigms 1940s-2000s'. *International Journal of Comparative Sociology* 47: 235-57.

Strabac, Z. and Listhaug, O. (2008). 'Anti-Muslim prejudice in Europe: A multilevel analysis of survey data from 30 countries'. *Social Science Research* 37: 268-86.

Winder, R. (2005). *Bloody foreigners: the story of immigration to Britain*. London: Abacus.

"Un hombre mata a su ex-mujer". The underrepresentation of men in violence-against-women news in contemporary Spanish media discourse: A CADS perspective.

José Santaemilia
Universitat de València
jose.santaemilia@uv.es

This paper stems from two key assumptions about media discourse on sensitive issues (such as violence against women – VAW): (i) that it contributes to the construction of social definitions of reality; and that (ii) that the language used to construct reality is highly evaluative and ideological (Hunston 2010). We live in a media saturated world, as we are constantly exposed to linguistic practices or rhetorical strategies that prove to be instrumental in talking about socially-relevant matters such as VAW. In particular, mass media are essential in defining key terms for day-to-day communication purposes. According to Bergvall (1999: 284), the mass media "impose images and constructions of behaviour", thus determining one's ethical and ideological attitude towards sensitive social issues.

In this paper I will explore the term *hombre* ('man' in Spanish) in a large corpus of VAW news from the digital versions of Spanish dailies such as *El País*, *El Mundo* and *ABC*. This includes all the news stories published over a large period of time (2004-2015), and constitutes a sub-corpus of GENTEXT-N, a large ad-hoc corpus compiled at the *Universitat de València* and revolving around contemporary gender and sexual (in)equality issues. A mixed methodological approach bringing together corpus techniques (word frequencies, collocations, and concordances –see Baker 2010) and critical analytical tools (CDA, news values –see Hoey 2005, Baker et al 2008, Baker & Levon 2015, Bednarek 2017) unveils meanings and associations that sometimes differ significantly from the values promoted by Spanish traditional lexicography. My aim, then, is to study and compare the media discursive representations of *hombre* vis-à-vis their lexicographic definitions.

In comparison with 'woman' (a term that is clearly foregrounded in contemporary Spanish media discourse on VAW –see Santaemilia & Maruenda 2016), 'man' is somewhat underrepresented and is often omitted. The *woman* vs *man* pair is fundamental in a large number of contemporary discourses, and their interrelated meanings have been fluctuating over the last decades, experiencing constant re-adjustments and re-definitions that are of utmost importance for public representations of sensitive issues. Traditional lexicography (e.g. *DRAE* 2014) still presents men and women in a more or less stereotypical light (Calero 1999, Bengoechea 2000, Forgas 2000, Lledó 2002), thus perpetuating asymmetries (Cameron 1998), prejudices (Hoey 1996) or sexist traces (Stoll 2003, Cabeza & Rodríguez 2013) that have repeatedly been denounced or condemned.

Therefore, attention has been paid to such traits as word frequencies, strongest collocates, concordance data, discourse prosodies, lexical priming, agency, critical or ideological values, positivity/negativity and other news values, etc. The focus will be on the set of contextually determined meanings and resources employed to assign certain ideological values to the presence/absence of men in VAW news stories, as well as to the actions or roles they perform. It is evident that, though blame is increasingly assigned to men in gender-based violence episodes, women routinely still continue to be –linguistically, rhetorically and ideologically- at its very centre. I believe that revising contemporary social or ideological definitions of key terms (*woman*, *man*, *violence*, *victim*, *aggression*, and many others) constitutes an essential step towards a greater understanding of the changing public discourses on socio-ideological issues (such as VAW) in the Western world.

References

Baker, P., Gabrielatos, C., KhosraviNik, M., Krzyżanowski, M., McEnery, T., and Wodak, R. 2008. "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to

examine discourses of refugees and asylum seekers in the UK press". *Discourse and Society 19*(3): 273-306.

Baker, P., Gabrielatos, C. and McEnery, T. 2013. *Discourse Analysis and Media Attitudes: The representation of Islam in the British press*. Cambridge: Cambridge University Press.

Baker, P. and Levon, E. 2015 "Picking the right cherries? A comparison of corpus-based and qualittive analyses of news articles about masculinity". *Discourse and Communication* 9(2): 221-236.

Bednarek, M. and Caple, H. (2017) *The Discourse of News Values. How News Organizations Create Newsworthiness*. Oxford: OUP.

Bengoechea, M. 2000. "En el umbral de un nuevo discurso periodístico sobre violencia y agencia femenina: de la crónica de sucesos a la reseña literaria". *Cuadernos de Información y Comunicación* 5: 9-22.

Bergvall, V. 1999 "Toward a comprehensive theory of language and gender". *Language and Society* **28: 273-293.**

Cabeza, M. and Rodríguez, S. 2013. "Aspectos ideológicos, gramaticales y léxicos del sexismo lingüístico". *Estudios Filológicos* 52: 7-27.

Calero, M. 1999 "Diccionario, pensamiento colectivo e ideología (o los peligros de definir)". In N. Vila., M. Calero., R. Mateu., M. Casanovas., and J. Orduña (eds.) *Así son los diccionarios*. Lleida: Universitat. 149-201.

Cameron, D. 1998. *The Feminist Critique of Language: A Reader*. London: Routledge. 2nd ed.

DRAE (*Diccionario de la Real Academia Española*). 2014. *Diccionario de la Lengua Española*. Madrid: RAE. 23rd ed. (online http://www.rae.es/)

Ehrlich, S. 2004. "Linguistic Discrimination and Violence against Women: Discursive Practices and Material Effects". In M. Bucholtz (ed.) *Language and Woman's Place: Text and Commentaries*. New York: Oxford University Press. 223-228.

Forgas Berdet, E. 2000. "La imagen social de la mujer en tres décadas de ejemplos lexicográficos". In F.J. Ruiz de Mendoza (eds.) *Panorama actual de la lingüística aplicada: Conocimiento, procesamiento y uso del lenguaje*. Vol. I. Logroño: Universidad de La Rioja. 1021-1030.

Hoey, M. 1996. "A clause-relational analysis of selected dictionary entries: contrast and compatibility in the definitions of 'man' and 'woman'". In C. Caldas-Coulthard and M.Coulthard (eds.) *Texts and Practices: Readings in Critical Discourse Analysis*. London/New York: Routledge. 150-165.

Hoey, M. 2005. *Lexical Priming; A new theory of words and language*. London: Routledge.

Hunston, S. 2010. *Corpus Approaches to Evaluation. Phraseology and Evaluative Language*. London: Routledge.

Lledó, E. 2002. "Crònica d'un equívoc: La construcció d'una identitat femenina en les notícies sobre maltractaments". *Lectora* 8: 87-97.

Santaemilia, J. and Maruenda, S. 2016. "The linguistic representation of gender violence in (written) media discourse: The term 'woman' in Spanish contemporary newspapers." In P. Bou-Franch (ed.) *Exploring Language Aggression against Women*. Amsterdam/Philadelphia: John Benjamins Publishing. 82-106.

Stoll, E. 2003. "Sexismo en el léxico: algunas observaciones acerca de las entradas mujer y hombre en diccionarios monolingües del italiano, español y francés". In J. Santaemilia (ed.) *Género, lenguaje y traducción*. Valencia: Universitat de València/Dirección General de la Mujer. 433-447.

Institutional discourse on same-sex marriage: a contrastive analysis between the US and the UK

Carmen Serena Santonocito
"Parthenope" University
cserena.santonocito@uniparthenope.it

Following Australian legalisation on same-sex marriages in December 2017, the institutional

debate leading to the same recognition for the US and for the UK has been in the spotlight once again. Moving from this recent debate, this study explores the representation of same-sex marriage in the institutional discourse of the UK and the US happening during the years which preceded and followed the ruling.

The theoretical framework is provided by seminal works by Partington (2009; 2010), Baker (2006) and Baker et al. (2008) where the methodological implementation of Corpus-Assisted Discourse Studies (henceforth, CADS) is applied. As a matter of fact, CADS has proved its consistency against many critical views on both corpus approaches and critical discourse studies. Its merit lies in allowing researchers to study discourse both from a qualitative and a quantitative perspective, thus enabling both approaches to complement each other and to overcome possible criticism that might be addressed to the single method. Indeed, in CADS discourse production is studied not only through the quantitative evidence deriving from the corpus, but also through the close reading of texts and in the broader context where discourse is embedded. The latter meaning to investigate also the relationship between texts, authors and the socio-political institution which produce them. This paper primarily focuses on a linguistic and discursive investigation, considering the assumption that language and discourse cannot be separated from the context in which they are produced.

Moving from institutional debate leading to a legal recognition of same-sex marriage both in the UK and in the US – i. e. the *Marriage (Same Sex Couples) Act 2013* for the UK, and US 2015 *Supreme Court ruling on the Obergefell vs. Hodges case* allowing same-sex marriage in the fifty States – this study aims to explore the discursive representation and the institutional reception of same-sex marriages in the official speeches, statements and press releases of the former political leaders in the UK ad in the US, i. e. Prime Minister David Cameron and President Barack Obama. In particular, the paper tries to answer the following research questions:

- How is same-sex marriage lexicalized in the institutional speech of each political leader?
- What kind of evaluation do the institutional interventions of Obama and Cameron convey for same-sex marriage?
- What are the similarities and differences in the discursive construal of same-sex marriage in the two countries?

As it can be easily inferred, the study starts with two parallel analyses which eventually converge in a contrastive analysis involving two English-speaking countries, therefore, two parallel ad-hoc corpora are built. The material is retrieved from the official institutional websites of both leaders. The ad-hoc corpora are of relatively small dimension because they address a specialized type of discourse with a limited quantity of data. Scholarly literature on the topic of specialized corpora (Mautner, 2009) claims that the size of 35,000/40,000 words can achieve a reasonable balance between the representativeness of data and the limited number of words available. The two corpora cover a slightly different time span, depending on the incumbency of the political leader. The Obama corpus convers from 2009 to 2017, while the Cameron corpus goes from 2010 to 2016. Following CADS methodology, a first quantitative search in the corpus is aimed to look for word frequencies and collocations. The analysis, gradually proceeding to a qualitative search, carries on with the close-reading of selected extracts and the individuation of the semantic prosody of word clusters connected with the representation of same-sex marriage

in order to elicit both lexicalization and evaluation. The criterion adopted to select the samples for the qualitative analysis is given by the statistical significance and the salience of lexical patterns used to represent same-sex marriage. Key elements considered in this phase are the main discursive strategies employed by discourse analysts; among them: the notion of modality (Fairclough, 1992), inclusive/exclusive discourse strategies and euphemisms (van Dijk, 2006), elements pertaining to the transitivity model – such as the use of agentless passive and nominalization (Mayr, 2008), agent description, authoritative argumentation and vagueness (Partington, 2006; 2010). In conclusion, the two parallel analyses converge into a comparative scrutiny highlighting similarities and differences between the two discursive representations.

References

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

Baker, Paul et al. (2008). "A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press". *Discourse Society*. 19(3): 273–306.

Fairclough, N. (1992). *Discourse and Social Change*. Cambridge: Polity Press.

Mautner, G. (2009). "Corpora and critical discourse analysis". In: P. Baker (Ed.). *Contemporary Corpus Linguistics*. London: Continuum. 32-46.

Mayr, A. (2008) (Ed.). *Language and Power*. London: Continuum.

Partington, A. (2006). *Persuasion in politics: a textbook*. Milano: LED.

Partington, A. (2009). "Evaluating evaluation and some concluding thoughts on CADS". In: Morley and Bayley (Eds.) *Corpus-assisted discourse studies on the Iraq conflict: Wording the war*. New York: Routledge. 261-304.

Partington, A. (2010). "Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: An overview of the project". *Corpora* 5 (2): 83-108.

van Dijk, T. A. (2006). "Politics, Ideology and Discourse". In: Brown K. (Ed.) *Encyclopaedia of language and linguistics*. Vol. 9. Amsterdam; Oxford: Elsevier.

van Dijk, Teun (2009). "Critical discourse studies: A sociocognitive approach". In: Ruth Wodak, Michael Meyer (eds.): *Methods for Critical Discourse Analysis*. 2nd ed. London: SAGE, 62–86.

Frugality and fairness. The 2008 economic crisis and morality

Rosa Escanes Sierra
University of Sheffield
r.escanes.sierra@sheffield.ac.uk

This paper contributes to the body of work that tries to understand the links between morality and evaluative language in elite political discourse within the context of the 2008 economic crisis in the UK (see for example Kelsey et al, 2016). Austerity measures were one of the main legacies of this major event and still resonate in British politics. The reduction of the state, legitimised by deficit panic tactics (Krugman, 2012) is still part of the political agenda.

The study, within a Critical Discourse Analysis framework, is part of a doctoral project which aims to enhance our knowledge regarding the role of ethics in the discourse of political economy. Specifically, it considers whether an economic crisis could increase the use of overt morality discourse and what exactly this explicitness looks like linguistically. It contemplates this question in terms of how intrinsic morality is within politics and economics. Ethics are part of the very purpose of government, since they determine duties for citizens and translate values

into policies (Swift, 2006). Moreover, economics, was actually born as a branch of ethics (Sen, 1987).

This research pays particular attention to the role of income inequality in moral linguistic patterns. Inequality was one of the main focuses when analysing the failure of the neoliberal model in 2008. 'Remoralising' the markets by acknowledging their problems with greed and polarised wealth distribution was, at the rhetorical level, a common trend amongst some elite politicians at that time (Jessop, 2012). In view of this, and considering their role in wealth distribution, I analyse yearly budget parliamentary debates in Britain from 2008 to 2012 and compare them to an equivalent corpus from a period of economic growth, 2002-2006. This comparison aims to see if there was indeed a more prominent use of morality invocations during the crisis period, using frequency and keyword analysis, and how this relates to the framing of inequality.

Furthermore, for this paper I focus on two concepts dealt with in the study: frugality and fairness. They serve as examples to discuss how corpus linguistics can be used to break down a complex idea like morality. Using tools such as keyword analysis, semantic domains, collocation patterns and close textual analysis, the study proposes an inductive dialogue between a top-down (from big concepts or categories to specific lexical items) and a bottom-up (from specific lexical items to big concepts or categories) perspective.

References

Jessop, B. 2012, November. "Recovered imaginaries, imagined recoveries: a cultural political economy of crisis construals and crisis-management in the North Atlantic financial crisis". Paper presented at the *Strategies of representing and managing crises* conference, Morecambe, UK.

Kelsey, D., Mueller, F., Whittle, A., & KhosraviNik, M. (2016). "Financial crisis and austerity: interdisciplinary concerns in critical discourse analysis". *Critical Discourse Studies, 13*(1).

Krugman, P. 2012, May 31. "The Austerity Agenda". The New York Times.

Sen, A. 1987. "On ethics & economics". Oxford: Blackwell.

Swift, A. (2006). "Political philosophy. A beginner's guide for students and politicians". (2nd, Ed.) Cambridge: Polity.

A Collocational Analysis of Synonymous Adverbs of Frequency in COCA: *Rarely, Seldom, Infrequently, Hardly, Scarcely,* and *Barely*

Sibel Söğüt

Anadolu University

sibelsogut@anadolu.edu.tr

Ümit Deniz Turan

Anadolu University

udturan@anadolu.edu.tr

Near-synonymous words are lexical pairs that have very similar cognitive or denotational meanings, but may differ in collocational or prosodic behaviour (Xiao & McEnery, 2006: 108), i.e. synonymous words are not necessarily collocationally interchangeable (Tognini-Bonelli 2001:34). This research study examines collocational properties of a set of six near-synonymous adverbs of frequency with a negative sense- *rarely, seldom, infrequently, hardly, scarcely* and *barely*. When dictionary definitions and synonymy entries are taken into consideration, these adverbs seem to be interchangeable. However, there are systematic differences in near-synonymous words' context of use, i.e. in naturally-occurring discourse (Arppe, Gilquin, Glynn, Hilpert, & Zeschel, 2010). Thus, a corpus analysis is necessary to illuminate the subtle or not so subtle differences among these adverbs. With this aim in mind, the present study investigates

six negative adverbs of frequency in 560 million words of text - *Corpus of Contemporary American English* (COCA). COCA was chosen as the corpus because it is the most recent comprehensive and widely used corpus consisting of five genres of spoken, fiction, popular magazines, newspapers, and academic journals in the period 1990-2017. The study implements an analysis model of identifying collocations using the first word to the right of the target adverb. In addition, if a collocation with a word on its left is very salient, this distinctive role is also reported. The Behavioral Profile Analysis Approach (Liu & Espino, 2012) was used and nouns, verbs, adjectives, and adverbs each adverb typically modifies were and the results were manually checked. The research questions addressed within the scope of the study are as follows: a) What is the frequency distribution of these adverbs depending on genres? b) What are the collocational differences of these adverbs?

The overall findings of the study show that these adverbs differ in terms of contextual uses across registers and that they have different collocational patterns. In terms of genres, *scarcely*, *barely* and *hardly* were found to be more frequent in fiction, *rarely* and *seldom* in magazine and academic discourse, *infrequently* in academic discourse. The most frequently used collocations in the category of verb collocations are *scarcely be, seldom be, rarely see, barely have, hardly be* and *infrequently use.* There are a number of differences in terms of the adverb-noun collocations, however, they are not significant or clear cut, the collocations are *scarcely time, seldom scene, rarely part, barely time, hardly news* and *infrequently suspects*. In terms of the adjective collocations, *scarcely visible, seldom, rarely, infrequently used, barely visible* and *hardly surprising* were used together whereas *hardly, rarely, scarcely ever, seldom, barely even* and *infrequently now* were frequent collocations. The overall findings revealed that some adverbs of frequency dominate in formal academic language; others are used in daily speech. For example, *only rarely* mostly occurs in academic discourse whereas it is quite rare in spoken language. Although expressions such as *rarely seen* and *not infrequently* were frequently used in COCA, *infrequently seen* or *not rarely,* and *not scarcely* would sound odd. In certain contexts, these adverbs are not used interchangeably as shown in the following example: *People were gawking at my father because he hardly ever went to parties, but most of them were keeping a respectful distance. (FIC) (We do not say 'he seldom ever went to parties'.)*

References

Arppe, A., Gilquin, G., Glynn, D., Hilpert, M., & Zeschel, A. (2010). Cognitive corpus linguistics: Five points of debate on current theory and methodology. *Corpora, 5*(1), 1-27.

Davies, M. (2010) Corpus of Contemporary American English (COCA). Retrieved from https://corpus.byu.edu/coca/

Liu, D. (2010). Is it a chief, main, major, primary, or principal concern?: A corpus-based behavioral profile study of the near-synonyms. *International Journal of Corpus Linguistics*, *15*(1), 56-87.

Liu, D., & Espino, M. (2012). Actually, Genuinely, Really, and Truly: A corpus-based Behavioral Profile study of near-synonymous adverbs. *International Journal of Corpus Linguistics, 17*(2), 198-228.

Tognini-Bonelli, E. (2001). *Corpus linguistics at work* (Vol. 6). John Benjamins Publishing.

Xiao, R., & McEnery, T. (2006). Collocation, semantic prosody, and near synonymy: A cross-linguistic perspective. *Applied linguistics, 27*(1), 103-129.

Combining keyness and dispersion statistics to investigate the representation of rape in the British press

Alessia Tranchese
University of Portsmouth
alessia.tranchese@port.ac.uk

Adriano Ferraresi
University of Bologna
adriano.ferraresi@unibo.it

Keyword analysis is arguably one of most widely used methods in corpus linguistics (McEnery/Hardie 2012:43). Possibly due to its centrality to the discipline, much attention has been devoted in recent years to developing more rigorous methods for keyness statistics calculation (e.g. Gabrielatos 2018), as well as for the qualitative analysis of keywords themselves (e.g. Marchi/Taylor 2009). Focusing on a corpus of British newspaper articles dealing with rape crimes, this paper pursues two aims. First, it explores a still underused approach to keyword analysis, one that combines keyness statistics and dispersion values (Gries 2008) to assess how evenly keywords are distributed over different corpus parts, and hence how "typical" (or not) they are of the corpus as a whole. Second, it proposes one of the first systematic investigations of how rape is represented in the British press, bringing together corpus methods and a Critical Discourse Analysis (CDA) perspective.

The representation of violence against women in the press has been the subject of scholarly interest before, especially within the Critical Discourse Analysis paradigm (Fairclough 1995). Notable examples include Soothill and Walby (1991) who carry out a content analysis of the representation of sexist crimes in the news, and Clark (1992) who analyses the way in which *The Sun* represents women in news reports of sexist violence. Yet, to the best of our knowledge, no previous study has adopted this quantitative/qualitative framework in an analysis of the representation of rape in the press.

The corpus itself was collected by sampling texts from the online database Nexis, considering four national British quality papers (*The Guardian*, *The Independent*, *The Times* and *The Daily Telegraph*), and using "rape* OR rapist* OR raping" as search terms. The resulting corpus includes 484 articles and 250,841 words.

In our study, keywords are obtained by comparing the research corpus with a closely matched reference corpus of articles published by the same newspapers and over the same time period. Keyword extraction is based on Log-Likelihood (Rayson/Garside 2000) and the %DIFF effect size measure (Gabrielatos/Marchi 2012). For each of the top 100 keywords, dispersion values, specifically deviation of proportions (DP; Gries 2008), are then calculated over different subcorpora – rather than individual texts (cf. Paquot/Bestgen 2009) – matching different time periods and different newspapers represented in the corpus. These 100 keywords are subjected to concordance and collocation analysis, considering their uniform vs. less uniform dispersion (DP < 0.33 vs. DP ≥ 0.33; cf. Gries 2008).

Results suggest that keywords with uniform dispersion point to linguistic choices reinforcing rape myths, and are suggestive of stereotypical attitudes towards gender roles (Benedict 1992). On the other hand, high DP values are found to be reliable indicators of keywords associated with especially "newsworthy" (Bednarek/Caple 2017) rape cases and specific newspapers. The paper concludes by discussing implications of these findings in the light of the current debate on objectivity in (news) discourse analysis.

References

Bednarek M. and H. Caple (2017). *The Discourse of News Values*. Oxford: Oxford University Press.

Benedict, H. (1992). *Virgin or Vamp: How the Press Covers Sex Crimes*. New York: Oxford University Press.

Clark, K. (1992). "Representations of Women in The *Sun's* Reporting of Crimes of Sexual Violence", in M.J. Toolan (ed.), *Language, Text and Context*. London: Routledge. pp. 183-197.

Fairclough, N. (1995) *Critical Discourse Analysis: The Critical Study of Language. London: Longman.*

Gabrielatos, C. (2018). "Keyness analysis: nature, metrics and techniques". In Taylor, C & Marchi, A. (eds.) *Corpus Approaches To Discourse*. Oxford: Routledge.

Gabrielatos, C. and A. Marchi. 2012. "Keyness: Appropriate metrics and practical issues". Paper presented at *Critical approaches to discourse studies 2012*, Bologna, 14 September 2012.

Gries, S. Th. (2008). "Dispersions and Adjusted Frequencies in Corpora". *International Journal of Corpus Linguistics*, 13(4): 403-437.

Marchi A. and C. Taylor (2009). "If on a Winter's Night Two Researchers… A Challenge to Assumptions of Soundness of Interpretation". *CADAAD Journal*, 3(1): 1-20.

McEnery, T. and A. Hardie (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.

Paquot, M. and Y. Bestgen (2009). "A Comparison of Three Statistical Tests for Keyword Extraction". In A. H. Jucker, D. Schreier and M. Hundt (eds.) *Corpora: Pragmatics and Discourse*, pp. 247-269. Leiden: Brill.

Rayson, P. and R. Garside (2000). "Comparing Corpora using Frequency Profiling". In *Proceedings of the Workshop on Comparing Corpora*, p.1-6. October 7, Hong Kong.

Soothill, K., Walby, S. (1991). *Sex Crime in the News*. London: Routledge.

Negative identifiers and nominal compounds. Conceptually profiling complex identifying noun phrases to critically examine their discursive functions in a corpus of web forum discussions

Eva Triebl
Lancaster University
evaetriebl@gmail.com

People regularly dissociate themselves from certain identity categories, saying what they *are not* rather than saying what they *are*, whether for actually rejecting a category or for purely strategic reasons. Examining negative identifiers expressed in structures of the type "*I* + COPULA + *not* + INDEFINITE NP" in a corpus allows us to take a closer look at concepts that people regularly dissociate themselves from.

While a focus on the meaning classes of the noun phrases and their quantitative relations seems the most obvious approach to this question, I will concentrate on their forms in this study. NPs in negative identification may be simple (e.g. *I'm not a fan*) or complex (e.g. *I'm not a PREMODIFIER fan POSTMODIFIER*), if complex, they may contain a nominal compound (*I'm not an Apple fan*), a premodifying adjective (e.g. *I'm not a hard-core fan*), a postmodifying prepositional phrase (e.g. *I'm not a fan of smartphones*), a relative clause (e.g. *I'm not a fan who has stopped thinking for herself*), a participle construction (e.g. *I'm not a fan trying to cause trouble*), or any combination of the former (e.g. *I'm not a great olive fan*). It might be argued that these different formal categories fulfil different functions depending on whether we are dealing with a simple NP or a complex NP, whereby the degree of compactness of the conceptual unit the latter represent can be considered to decrease the further we move from compound, representing the tightest conceptual connection, to postmodification, constructing weaker conceptual links between modifier and head noun and, thus, weaker cases of negative identification: for example, negatively identifying with the compound *Apple fan* entirely negates identification with the concept of *fan*, whereas negatively identifying as *fan trying to cause trouble* only serves to contrast the speaker with a particular kind of fan rather than with the entire conceptual category.

This study seeks to explore how simple and (the various classes of) complex identifying NPs are distributed (quantification of formal structures), how these different formal categories relate to semantic ones (cross-categorization with semantic categories) and what implications these patterns of usage have. In other words, I want to examine if particular conceptual categories of identifying NPs can be related to particular formal structures in patterned ways, i.e. if conceptual compactness can be found to correlate more with certain semantic categories than with others, constructing particular identity categories as conceptually more compact and thus more central, or permanent aspects of speakers' identities. It is, for instance, theoretically possible that identifying NPs in disclaimers of expertise – *I'm not an […] expert […]* – tend to be more frequently expressed by compounds, constructing a more compact relation between the associated concepts (e.g. *I'm not a Java expert* rather than *I'm not an expert in Java*) than NPs denoting preferences (e.g. *I'm not a fan of diesels* rather than *I'm not a diesel fan*), or do formal and semantic categories not correlate.

Drawing on the theoretical concepts of corpus-based CDA, this study will qualitatively and quantitatively examine this question in a self-compiled, randomly sampled corpus of 940 negative identifiers used in UK web discussion forums. It will show according to which criteria a formal-functional framework of negative identification can systematically be set up and methodologically operationalized in corpus analysis aiming to identify transtexual patterns of language use, bridging the gap between microlinguistic analysis of paradigmatically defined linguistic elements and contextually dependent functions of negative identifiers to address questions in socially-oriented research on language and identity.

Right- and left-wing identities: a corpus-assisted study of Hungarian parliamentary debates

Elena Valvason
University of Pavia/University of Bergamo
elena.valvason01@universitadipavia.it

Between 1998 and 2003 Hungary negotiated its accession to the European Union. Opinions and attitudes towards EU-membership varied across the country. Also the concerns of the Hungarian parties in relation to the EU differed significantly and they resulted in heated debates. Two major political forces emerged during these discussions: Fidesz (the 'Hungarian Civic Alliance'; a right-wing, liberal party) and MSZP (the 'Hungarian Socialist Party'; a left-wing, socialist party). Fidesz and MSZP sided with opposed ideas and they neared the negotiations to join the European Union with different agendas. They debated on diverse matters whilst moulding diverging identities (Batory 2008, Taggart and Szczerbiak 2013).

In the current paper I aim to describe the linguistic identity that Fidesz and MSZP created for themselves when they discussed about the EU during parliamentary debates, in terms of who they claimed to be and what they sustained to do. The approach to the study is corpus-assisted, the methodology both quantitative and qualitative (cf. Partington et al. 2013).

I define Fidesz's and MSZP's identities by examining two sub-corpora of the HUNPOL corpus. I collected the HUNPOL corpus including Hungarian parliamentary debates about the European Union produced between 1998 and 2003. The sub-corpus containing speeches uttered by Fidesz

representatives is called HUNPOL-Fidesz and it is a collection of 137 texts (97,722 running tokens). The one including speeches delivered by MSZP delegates is HUNPOL-MSZP and it consists of 102 texts (71,467 tokens).

For the analysis, I first outline the characterization of the parties as political beings by observing how they referred to themselves during the aforementioned parliamentary debates. I analyse the *concordances* of the names of the two parties and of the personal pronouns *mi* 'we' and *minket* 'us'. Concordances, in fact, emphasise the co-text that surrounds a word in all its occurrences and they provide details of the term's linguistic behaviour (e.g. Sinclair 1991). For Fidesz I employ the expressions *Fidesz** and *Magyar Polgári Párt** ('Hungarian Civic Party*'; the second name of the Fidesz party). For MSZP I focus on the concordance lines of *MSZP** and of *Magyar Szocialista Párt** ('Hungarian Socialist Party*'; the extended form of MSZP). I expand on the concordance lines of the words through AntConc's Concordance Tool (Anthony 2018). I search for the collocates of the terms using a wildcard *, in order to assure that all the inflected forms of the words are included in the corpus (e.g. *MSZP-nek* 'to MSZP').

Secondly, I describe the identity of Fidesz and MSZP as political actors by sketching the main issues that the parties dealt with when discussing about the EU. To do so, I profit from the concept of *keyword*. Keywords statistically highlight the most peculiar linguistic features of a corpus in comparison with another: they are signposts to the "aboutness" of a corpus (Scott and Tribble 2006: 59; see also Bondi and Scott 2010). In the current research, they allow to signal the themes that the politicians addressed most frequently in their speeches. I extract a list of keywords for each party and then I peruse the concordance lines of the most significant keywords. I build the keyword lists with the Keyword List Tool of AntConc, utilising Log-likelihood as association measure and 15.13 as minimum keyness score. I use the two sub-corpora as node and reference corpora in turn, in order to highlight the most salient words in the two collections.

Results from this analysis show that in the concordance lines Fidesz described themselves as active political beings (that *támogatja* 'supports', that *segíteni kíván* 'wishes to help'), engaged in creating a political network across the country, in protecting the needs of the weakest, and in supporting enterprises like Hungary's joining the European Union. MSZP depicted themselves as honest politicians who fought with a clear idea in mind (e.g. *A Magyar Szocialista Párt fontosnak tartja, hogy […]* 'The Hungarian Socialist Party deems it important that […]' and *mi döntsük el […]* 'we decide that […]'). At the same time, the keywords in Fidesz's speeches (42 keywords) show that the party paid much attention to financial, political, and legal issues (with words like *adó* 'tax', *polgármester* 'mayor', and *törvény* 'law'). The keywords in HUNPOL-MSZP (41 keywords), on the other hand, stress that MSZP was mainly concerned with employment (e.g. *középvállalatok* 'medium-size companies', *munkanélküli* 'unemployed') and agriculture (as in *vidékfejlesztési* 'of (the) development of rural areas', *mezőgazdasági* 'agricultural').

References

Anthony, L. 2018. *AntConc (Version 3.5.7) [Computer Software]*. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/software.

Batory, A. 2008. "Euroscepticism in the Hungarian Party System: Voices from the Wilderness?" In A. Szczerbiak and P. Taggart (eds.) *Opposing Europe? The Comparative Party Politics of Euroscepticism*. Vol. 1 (pp. 263-276). Oxford: Oxford University Press.

Bondi, M., and Scott, M. (eds). 2010. *Keyness in Texts*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Partington, A., Duguid, A., and Taylor, C. 2013. *Patterns and Meanings in Discourse. Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Scott, M., and Tribble, C. 2006. *Textual Patterns: Key words and corpus analysis in language education*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Taggart, P., and Szczerbiak, A. 2013. "Coming in from the Cold? Euroscepticism, Government Participation and Party Positions on Europe". *Journal of Common Market Studies, 51*(1), 17-37.

Constructing Trust by Reducing Complexity: Fiscal Policy in the UK Press, 2010-2016

Catherine Walsh
Cardiff University
Walshc9@cardiff.ac.uk

How does the UK press construct fiscal analysis as trustworthy? Fiscal-policy analysis-units like the Institute for Fiscal Studies (IFS), National Audit Office (NAO) and Office for Budget Responsibility (OBR) are vocal contributors to the construction of economic knowledge in British political life.

Through the press, they can act as critics and/or vouchsafes for the Government's stewardship of the UK economy. But the policy judgments of these units are only effective insofar as they are trusted by other economic and political elites, and by politically-engaged citizens. Drawing on trust scholarship from sociology, politics, and organizational studies – and especially Niklas Luhmann's *Vertrauen* (1968/2017) – this study investigates discursive constructions of such institutional trust in the press. Representations of these three units were found in approximately ten-thousand articles published in the *Financial Times, Independent, Guardian, Telegraph* and *Times* between May 2010 and December 2016. This corpus was analysed with the programme *Antconc 3.5.0*, yielding concordances, clusters, collocates, and keyness measures made via comparisons to the British English 2006 (BE06) corpus. From a discursive perspective, indications of trustworthiness associated are of particular interest: accountability, authoritativeness, competency, expertise, familiarity, independence, morality, being rule-bound, self-confidence, transparency, and being already trusted by others. Similarly, from a Luhmannian perspective, indications of certainty and uncertainty were of interest: prediction, estimation, forecasts, warnings, confirmations, and the use of modal verbs.

When compared to the BE06, the corpus as a whole indicates pre-occupation with time, especially the future, which is a key context for constructions of trust. The modal verbs *will* and *would* are conspicuously well-represented, suggesting preoccupations with judgments of likelihood, with inevitability, expectation, probability, capability, or habit. At the same time, negative keyness in comparison to BE06 indicates a distance in the corpus from certainty and claims of knowledge or understanding. In proximity to the named units (OBR, NOA, and IFS), the results do demonstrate a wide range of explicit and implicit constructions of trustworthy attributes, but, surprisingly, the construction of their authority, competency, and expertise is not particularly common. This finding confounds a longstanding academic literature on the

importance of expertise in the construction of economic knowledge. The investigation also found that discursive tensions exist in the texts between constructions of dependence on vs. independence from Government. The tension evident in dependence vs. independence is interpreted as a trade-off between the authority of Government and a lack of transparency, which respectively build and dismantle institutional trust. Most importantly, the results show that the reduction of uncertainty through prediction and estimation is more pervasive and consistent than other features. The most common constructions of analysis-unit behaviour are forecasting, estimating, predicting, expecting, projecting, and revising. These constructions of fiscal analysis build trust by implying a decreased uncertainty while carefully maintaining distance from any absolute guarantes of future behaviour. These press findings support a Luhmannian interpretation of trust as a necessary and necessarily communicative act, one that reduces the perception of uncertainty about an inherently unknowable future. The construction of policy-analysis units is interpreted as simplifying an unacceptable economic and fiscal complexity by substituting acceptable expectations of the future behaviour of these units, the Government, and the economy.

References

Barbalet, J. 2009. "A characterization of trust, and its consequences". *Theory and Society*, 38, 367-382.
Luhmann, N. 2017. *Trust and power*. Cambridge: Polity Press.
Möllering, G. 2005. "The trust/control duality: an integrative perspective on the positive expectations of others". *International Sociology*. 20, 283-305.

Examining the Case for Old Bailey Trial Testimony as Linguistic Evidence

Julie Weeds
University of Sussex
J.E.Weeds@sussex.ac.uk

Fraser Dallachy
University of Glasgow
fraser.dallachy@glasgow.ac.uk

Justyna Robinson
University of Sussex
Justyna.Robinson@sussex.ac.uk

Tim Hitchcock
University of Sussex
T.Hitchcock@sussex.ac.uk

The proceedings of the Old Bailey, as captured in the Old Bailey Corpus (Huber et al. 2012), are a potentially rich source of accurately recorded historical speech. The project captured data from almost 200,000 trials over a 250year period, totalling ca. 134 million words of transcribed speech together with socio-biographical and pragmatic about the speaker and their role in the courtroom. So far, however, historians still debate (Leguy, 2017) whether the court proceedings are true representations of speech; since the role of the court reporter could have involved a fair amount of censorship. This paper uses methods from corpus and computational linguistics to investigate whether the recorded speech incorporated in the Old Bailey Corpus reflect real voices – that the language recorded is distinctive enough to illustrate that the shorthand recorder was producing a 'transcript' – however imperfect – rather than a digest synopsis of the trial testimony. This is done through a case study investigating the social dynamics between members of different sections of society by examining their language within the 19th century courtroom.

More specifically, in this case study, we investigate language variation in the 19th Century

courtroom according to gender and role. We restrict our study to a 20-year time-frame and trials for theft in order to limit variation caused by these other factors. Specifically, we look at the ways in which men and women were described by themselves and others. We look at grammatical structures within the language since they give access to understanding relations of power. For example, the choice of active or passive voice informs us as to who is allowed to perform actions and who are the passive recipients of actions (in the mind of the speaker). We also assert that the grammatical structures within the language are less susceptible to conscious modification. For example, it is highly unlikely that court reporters would change active voice to passive voice (or vice versa) consistently for one gender but not the other.

First, we query this large historical dataset (approx. 10 million tokens) at the level of concept, since a single concept can be denoted by different words and a single word can be used to denote different concepts. We use the SAMUELS semantic tagger (Piao et al. 2014), which is sensitive to historical word use, to carry out automatic conceptual tagging. Second, the distributions of concepts are modelled in terms of the co-occurring concepts and grammatical relations between concepts. We use SpaCy (Honnibal and Johnson, 2015) to annotate the text in terms of grammatical dependency relations and then extract triples which represent co-occurrences of concepts and grammatical relations. Third, we use localised positive pointwise mutual information (PPMI) to rank co-occurrences (Evert, 2005). These rankings provide characterisations of concepts (and grammatical relations) according to the corpus or sub-corpus from which they are derived.

We present analysis of a number of different concepts and their corresponding characterisations in the speech of males and females (according to court role) in this dataset. In particular, we show that the adjectives chosen by women to describe women are different to the adjectives chosen by women to describe men, which in turn are more similar to the adjectives chosen by men to describe both men and women. For example, women more frequently chose adjectives relating to social or economic status such as "poor", "single" and "married" when referring to women than when they were referring to men and also more frequently than men did when referring to either gender. In conclusion, we are able to show significant differences in the conceptualisation of gender by people of different genders (but having the same role in the courtroom) in this historic dataset. This evidence provides support for the use of these transcripts of court proceedings as true representations of historical speech.

References

Evert, S. "The Statistics of Word Co-occurrences: Word Pairs and Collocations". PhD thesis, Uni. of Stuttgart, 2005

Honnibal, M. and M. Johnson. "An improved non-monotonic transition system for Dependency Parsing." *EMNLP 2015*

Huber, M., et al. "The Old Bailey Corpus. Spoken English in the 18th and 19th Centuries". www.uni-giessen.de/oldbaileycorpus. 2012 (accessed January 2018)

Leguy, I. "Women's Voice: Interaction and Performance in the Courtroom, 1800-1810". *Women's History, 2(8)*. 2017

Piao, S., et al. "Developing the Historical Thesaurus Semantic Tagger". *DHC 2014*

'Bisexual oysters': Challenges in locating marginal identities in a historical newspaper corpus

Mark Wilkinson
Lancaster University
m.wilkinson7@lancaster.ac.uk

The social position of *lesbian*, *gay*, *bisexual*, *transgender*, *queer*, and *intersex* (LGBTQI) people in Britain has changed dramatically since 1957 when The Wolfenden Report first made recommendations for the partial decriminalisation of homosexuality. In addition to significant legal changes over the past 70 years, there has also been a cultural shift in which queer identities have transitioned from a marginal position in the past to a more broadly accepted position in society today (Weeks, 2017). Using corpus data can assist in mapping how such cultural changes have been both reflected in and constituted by language. The challenge however, is how to locate such marginal identities in historical corpora. As stated by Baker and McEnery (2017), the very nature of marginal identities means that the frequency of lexis referring to such identities will likely be suppressed.

This paper therefore discusses some of the challenges encountered in trying to build a corpus based on the representation of LGBTQI people and communities between 1957-2017. Due to its longevity as a publication and its position as a centrist broadsheet, it was decided that *The Times* would provide an appropriate data set. Unlike a right-wing tabloid or a more left-leaning broadsheet, *The Times* would not likely take any positions perceived to be radical in their discussion of queer identities. Rather, *The Times* is more likely to represent a consensus position on how LGBTQI people and communities are perceived generally and, therefore, perceived by the majority of Britons.

Deciding which key terms to use in the building of a historical newspaper corpus is a significant challenge when attempting to represent how LGBTQI people have been discursively constructed through language. In addition to an absence of lexical items explicitly discussing queer identities and practices, the corpus linguist must also account for euphemistic terms and unfamiliar slang. This paper, however, focuses on the issue of diachronic variation in lexical meaning when selecting for key terms to include in the corpus. Through a discussion of the term *bisexual*, it will be shown how identity and the language of identity are not only historically contingent, but that words themselves are dynamic and subject to rapid change within a matter of years. In the case of *bisexual*, the term was chosen to include in the corpus as the contemporary definition describes someone who is sexually attracted to members of both sexes (Oxford English Dictionary [OED], 2018). A closer examination of the word in context, however, quickly revealed that between 1957-1967, *bisexual* more frequently referred to someone or some*thing*, e.g. oysters, that had both male and female reproductive organs and capabilities – an identity that would now likely be referred to as intersexual (OED, 2018). In fact, only 43% of the occurrences of *bisexual* referred to someone who was sexually and romantically attracted to members of both genders. A corpus of the term bisexual was then built in order to track the development of the term, demonstrating that, up until 1989, *bisexual* – in addition to referring to intersexual organisms – was also often used to refer to professions, fashions, sports, and names that were available for members of both genders. While the contemporary meaning of bisexual was present in the data, a close reading of concordance lines revealed that it did not become the majority use until the mid- to late 1980s.

If the goal of the study is to reveal how the changing position of LGBTQI people and communities has been reflected in and constituted by language, we first have to establish *what* the language of queer identity is. As is demonstrated by the word *bisexual*, identity labels cannot always be read backwards in time. On the contrary, uncovering the queer past requires a careful analysis of lexical items in their historical context and an analysis that accounts for the historical contingency of identity – both of the subject and the researcher.

References

McEnery, T. and Baker, H. 2017. "The public representation of homosexual men in seventeenth-century England – a corpus based view". *Journal of Historical Sociolinguistics.* 3(2): 197–217

Oxford English Dictionary. [Online] 2018. [Accessed 14/01/2018] Available from:
http://www.oed.com/view/Entry/98315?redirectedFrom=intersexual#eid

Oxford English Dictionary. [Online] 2018. [Accessed 14/01/2018] Available from:
http://www.oed.com/view/Entry/19448?redirectedFrom=bisexual#eid

Weeks, J. 2016. *Coming Out: The Emergence of LGBT Identities in Britain from the Nineteenth Century to the Present.*

Can letter frequencies identify contextual dimensions of discourse? A resampling experiment based on BNC data

Andrew Wilson
Lancaster University
a.wilson@lancaster.ac.uk

Establishing common ground is important for enhancing communication success. "Common ground" can extend beyond the superficial alignment of style in real-time dialogues to the alignment of ideological "angles" in texts written by an author for a target audience. In previous research, such a commonality of ideological *context* has been assessed using multivariate analysis of word or semantic category frequencies, leading to four stable contextual dimensions of "normative", "pragmatic", "emotional", and "rational" (Cleveland, 1971; McTavish & Pirro, 1990). New texts can be scored on these dimensions in a similar way to Biber's (1991) stylistic dimensions. However, this requires access to custom-designed lexical tools, and the success of the method can still be much affected by lexicon coverage, etc. A simpler, more general method would thus be welcome. This poster will investigate how far statistics based purely on letter frequencies can provide comparable results.

Letter counting was one of the first things to be done with the first generation of computer corpora (Zettersten, 1969), but it appears to have been largely neglected in more recent corpus linguistics, despite quite extensive recent work in mathematical linguistics (of which Pande & Dhami (2009) is but one example) and successful applications to areas such as authorship attribution (e.g. Grieve, 2007). Examining simple letter counts across corpus genres suggests that there may well be enough systematic variation for them to have a role in scoring texts on contextual dimensions.

However, the "onceness" of the reference corpus is a sticking point. We cannot be sure how far the frequencies and rankings would change if different texts had been sampled, but new genre-diverse corpora are rarely built until older ones become obsolete – cf. the twenty year gap between the BNC and BNC2014. This poster will tackle the problem by using an empirical

resampling approach to study the variation we might expect to see in letter counts and their rankings. Using a subset of texts from the BNC, selected to mirror closely the contextual dimensions of Cleveland (1971) and McTavish & Pirro (1990), multiple random samples of *N* running characters each will be extracted. This will allow empirical confidence intervals for the mean or median letter-frequency values to be established. (It has been observed that text samples as small as 1,000 running characters are sufficient for studying letter counts – Kelih, 2006.) The approach is somewhat akin to the bootstrap method in statistics (Efron & Tibshirani, 1993), but, here, multiple samples of running text are drawn directly from the corpus, rather than resampling from a single set of figures.

Using the results obtained, we will be able to establish how far letter-frequency statistics can be used for scoring text samples on these contextual dimensions.

References

Biber, D. (1991). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Cleveland, C.E. (1971). *Institutional rhetorics as universes of discourse*. Ph.D. dissertation, Northwestern University, Evanston, IL.

Efron, B. & Tibshirani, R.J. (1993). *An introduction to the bootstrap*. Boca Raton: Chapman & Hall/CRC.

Grieve, J. 2007. Quantitative authorship attribution: An evaluation of techniques. *Literary and Linguistic Computing*, 22, 251-270.

Kelih, E. 2006. "Häufigkeiten von Graphemen und Lauten: Zwei Ebenen - ein Modell? (Re-Analyse einer Untersuchung von A.M. Peškovskij)". In: P. Grzybek & R. Köhler (eds.), *Exact methods in the study of language and text*. Berlin: Mouton de Gruyter.

McTavish, D.G. & Pirro, E.B. (1990). Contextual content analysis. *Quality and Quantity,* 24, 245-265.

Pande, H. & Dhami, H.S. 2009. Generation of a model for grapheme frequencies and its refinement and validation by group theoretic aspects. *Journal of Quantitative Linguistics*, 16(4), 307-326.

Zettersten, A. 1969. *A statistical study of the graphic system of present-day American English*. Lund: Studentlitteratur.

'Straight guy looking for a hung wank buddy': The construction of 'authentic' heterosexuality in men-for-men casual encounters ads

Łukasz Woźniakowski
Adam Mickiewicz University in Poznań, Poland
lwozniakowski@wa.amu.edu.pl

Does identifying as a straight man preclude pursuing same-sex sexual relations? It would seem so, but dozens of personal ads posted by self-identified straight men seeking sexual encounters with men on Craigslist seem to prove otherwise. Not only does this blur the boundaries between seemingly distinct social categories, but it also exposes the tension between sexual identity and erotic desire. In fact, negotiating one's straight identity while enacting same-sex desire is what creates a precarious and fascinating site of identity work. As proposed by Bucholtz and Hall (2004), sexual identity is the result of *intersubjectively* negotiated practices and ideologies. In order to attain and maintain their straight identity in the context of men-for-men personal ads, posters draw on other social categories and ideologies that index 'authentic' heterosexuality. This complex identity work highlights the fact that sexual and other identities are, as referred to by Bucholtz and Hall (2004: 494), inherently relational, unstable, contingent and contextually specific. Although a number of studies have looked at the discursive construction of heterosexuality and homosocial desire (e.g., Kiesling, 2005, 2006) and, in particular, the

construction of 'authentic' heterosexuality in the context of same-sex personal ads (e.g., Ward, 2008), the methodology employed was mainly qualitative.

This study, which is a work in progress, investigates the construction of 'authentic' heterosexuality in personal ads posted by self-identified straight men seeking sex with men from the 'casual encounters' section of the bulletin board website Craigslist (craigslist.org). Methodologically, the study draws on corpus linguistic methods such as keyword, collocation and concordance analysis and on Bucholtz and Hall's (2004) *tactics of intersubjectivity* analytical model. A specialised corpus was built comprising 30,000 ads posted within a 3-month period by men seeking men in the 'casual encounters' section of Craigslist – London. This corpus was then divided into two subcorpora – one comprising ads posted by men who explicitly self-identified as straight and another by those who did not. The corpus-based analysis was then combined with and informed by the *tactics of intersubjectivity* model with the focus on the dimensions of sameness versus difference and genuineness versus artifice.

The preliminary findings suggest that self-identified straight men authenticate their heterosexuality by: (1) establishing their *sameness* with other straight men by drawing on the concept of male camaraderie, constructing themselves as ordinary, average guys and by referencing their heterosexual relationships; (2) asserting the *genuineness* of their straight identity by emphasising their lack of experience in same-sex sex, constructing their desire as only a fantasy or scenario they want to act out and constructing their desired sexual practices as 'soft' such as, for example, *handjob* or *stroke*; (3) disavowing themselves from the gay community by using fewer gay slang terms and, at the same time, using more pejorative terms with reference to gay men.

References

Bucholtz, M., & Hall, K. 2004. Theorizing identity in language and sexuality research. *Language in Society, 33*(4), 469-515.

Kiesling, S. F. 2005. Homosocial desire in men's talk: Balancing and re-creating cultural discourses of masculinity. *Language in Society, 34*(5), 695-726.

Kiesling, S. F. 2006. "Playing the Straight Man: Displaying and Maintaining Male Heterosexuality in Discourse". In D. Cameron & D. Kulick (Eds.), *The Language and Sexuality Reader* (pp. 118-131): Routledge.

Ward, J. 2008. Dude-Sex: White Masculinities and Authentic' Heterosexuality Among Dudes Who Have Sex with Dudes. *Sexualities, 11*(4), 414-434.

A corpus-assisted diachronic study on the attitudes of British Broadsheet Newspapers towards Hong Kong's handover to China

Shentao Xie
Lancaster University
s.xie2@lancaster.ac.uk

Although the broadsheet newspapers in the UK and the political event of Hong Kong's reversion to China have been studied by many researchers, the attitudes of UK broadsheet newspapers towards this event are not fully investigated. This paper discusses two questions: 1) For UK broadsheet newspapers, what are the mechanisms of the attitude generation and presentation towards a political event? 2) Did their attitudes on this event change over time?

This study acknowledges the wide spread phenomenon of presenting bias and inclination in news reporting. There are multiple factors accounting for this. Firstly, news articles are written by human beings and the subjective construal of the world is unavoidable for a person however hard he tries to be detached or objective. Secondly, newspaper in the UK are run by private companies that make profits to stay in existence so that contents favoured by their target readers are more likely to be produced and presented. Social groups have their own interests and 'tastes'. Therefore, the attitudes of that social group over a certain political event might be projected to the newspapers they read.

This paper discusses the linguistic techniques that are used to present attitudes in a news article. The study shows that there are varied ways of presenting attitudes in a news article. Those linguistic techniques and devices can be grouped into three aspects: 1) syntactic; 2) semantic; 3) morphological; 4) lexical. The author built two corpora for extracting and analysing the patterns of language use: one contained news texts of 7 UK broadsheet newspapers from 01st January 1993 to 30th June 1997 and the other from 1st July 1997 to 28th December 2003. The news texts were extracted with the four keywords 'Hong Kong', 'China', 'handover' and 'takeover'. By looking into the concordance lines that contained them the author identified and discussed the linguistic devices and techniques that reveal the author's attitudes.

A political event might happen in a very short period but its influences might last long and that political events might contain and/or trigger a series of 'sub-events' that impose significant impact over time. For instance, the core political event – Hong Kong's handover to China – consists of a series of events including the multi-round negotiations, talks, the official sovereignty transfer ceremony and social and political (e.g. legislative, administrative, judiciary, etc.) changes after the ceremony. This study investigates from a diachronic perspective the entire series of events over time and the author discovered that British broadsheet newspapers' attitudes towards it are generally negative but the level of negation altered and that the explicitness of such attitude was weaker after the transfer of sovereignty.

The results indicate that UK broadsheet newspapers, in spite of the golden rule of being objective, do present a certain attitude towards this political event. Also, many forms of linguistic devices are employed with multiple strategies in expressing such attitudes. Moreover, newspapers' attitudes towards a single political event change over time.

Categorizing "the common folk" and "big sister" in Chinese online discussion: A corpus-based discourse analysis

Bingjuan Xiong
University of Colorado
Bingjuan.xiong@colorado.edu

The study adopts corpus linguistics tools to explore salient discursive moves used by Chinese Internet users to engage in online debates pertaining to government officials' conduct and law enforcement in China. The corpus used in this study consists of online discussions of two

contentious social events from Chinese social media sites. [21] We used a commercial web scraper WebHarvy (https://www.webharvy.com/) to automatically extract online posts related to these two events. Approximately 80,000 posts were collected. These posts comprise a 2-million word corpus, which was segmented using the Stanford Word Segmenter.

A keyword analysis showed that, not surprisingly, that discourse pronouns (e.g., 你/you, and 我/I) and topic related words (e.g., 城管/city inspectors, 房子/house) were distinctive feature of the posts. More surprising was the use of membership category terms such as 老百姓/the common folk and 大姐/big sister (and related terms 女/female, 女人/woman, and 阿姨/auntie) that did not simply index the protagonists appearing in these two events as the use of 局长/director did, but more importantly signified the online commenters' different interpretations of what it means to be a member of these social categories and of their ability to create social change.

For example, looking into the concordance lines and the collocates of the term 老百姓/the common folk, we find salient discursive patterns related to how Chinese online commenters categorize their citizenry as "the common folk" rather than the more neutral term "citizen/公民". The most salient characterization of the common folk is that they were portrayed as passive recipients of actions coming from different parties in Chinese society. Most frequently, they were the recipient of negative actions from the city inspectors and government officials. Action verbs and Chinese prepositions were used to depict the negative (oftentimes harmful and oppressive) actions that were imposed on the common folks. For instance, 对付/tackle (freq L: 597), 欺负/bully (freq L: 308), 出动/dispatch (freq R: 159), 逼/force (freq L: 43), and 欺压/oppress (freq L: 30) and prepositions 对/dui, 在/zai, 把/ba, and 让/rang. It is also notable that the common folk were also positioned as the recipients of help, protection, and support from the government, lawyers, and officials when they were treated unjustly. In other words, the common folks were mostly talked about as people who did not have the agency to fight for their own justice but rather had their rights and justice at the mercy of the government, officials, and other professionals (e.g., lawyers and journalists) who are supposed to voice their concerns and bring them justice.

Another very notable characterization of the common folks concerns the use of particular kinds of descriptors to depict these group as poor, marginalized, and ordinary, living at the bottom of Chinese society. Many commenters singled out "one/一个common folk" in their comments to highlight the contrast of a group of twenty some city inspectors when they confronted this Chinese woman (a single common folk) in front of her house. This contrast signifies not only this woman being out-numbered in this incident, but more importantly functions to depict the city inspectors as being unreasonably bullying this one common folk who was by herself and

---

[21] The first event was about a Chinese civilian arguing with city inspectors in front of her house regarding whether she had the right to put flower stands outside in the street in a popular tourist city Xiamen. The second event relates to the online exposure of a Chinese official's luxurious life style in order to reveal his links with corruption. Both social events went viral on the Chinese Internet (the first event appeared as a video recording posted on China's video-sharing site www.youku.com and the second appeared on China's popular blogging site www.tianya.com).

powerless. Similarly, commenters used descriptors such as 一般/ordinary, 小/petty, 穷苦/impoverished, 平民/civilian, 底层/the bottom to characterize the common folks as this powerless and disadvantaged group of people. This characterization sheds light on why online commenters typically portrayed the common folk (with contested boundaries about who should be included in this group) as the passive recipient of actions rather than active agents initiating social and political change for themselves.

A Contrastive Analysis of Presidents' Messages in Higher Education

Liu Ying
Caritas Institute of Higher Education, Hong Kong
liuying1213@gmail.com

The issue of how the trend of marketization has influenced the discursive practices in higher education has received increasing attention of discourse researchers. Contemporary universities have been observed to sell their programs to the global market and form the new pattern of promoting selling points in their external communication with the public (Fairclough, 1993). Previous studies have explored diverse genres, for example, mission statement (Connell & Galasiński, 1998), international student prospectus (Askehave, 2007), university brochures (Osman, 2008), advertisements for academic posts (Xiong, 2012), university websites (Zhang & O'Halloran, 2013) and president's speech at graduation ceremony (Han, 2014). Compared with these genres, the president's message is a relatively under-researched genre in the discourse of higher education. But it is worthy of more investigation because it is an essential marketing tool to build the image of the university and serves as a window to the university for prospective students, their parents and potential academics.

The present study adopts a genre-based approach to analyze how marketization and increasing competition has shaped the discursive practices of Chinese and American universities in their presidents' messages. The corpus is composed of 112 presidents' messages collected from the webpages of 224 top-tier universities in China and America. The research questions that we intend to address are the following: (1) Are there any differences in the rhetorical moves in the presidents' messages in Chinese and American universities? (2) Are there any differences in the discourse strategies in the presidents' messages in Chinese and American universities?

It has been found that the marketization of higher education has greatly influenced the university discourse. Due to different histories and cultures, it is natural that Chinese and American universities do not respond to this trend in the same way and have different focus in their selling and promoting points and strategies, as evidenced in the discourse of presidents' messages. Six core rhetorical moves have been identified in both Chinese and American university presidents' messages and four moves are shared but two moves are different. Results indicate that there are more social discourses in the presidents' messages in American universities and more interactive and personal tone is used in American universities but more authoritative tone is used in Chinese universities. Among the different moves, the moves to emphasize history, achievement and status in China play a crucial role in promoting Chinese universities whilst the moves of foregrounding defining features of the university and qualities of its people are important in the promotion of American universities. Furthermore, promoting academic quality is the most obvious strategy for Chinese universities to attract potential

applicants, but the core values such as individuality, creativity and innovation are highlighted to attract talented students and academics in American universities.

References

Askehave, I. 2007. "The impact of marketization on higher education genres – the international student prospectus as a case in point". *Discourse Studies*, 9(6): 723–742.

Connell, I. and Galasiński, D. 1998. "Academic mission statement: An exercise in negotiation". *Discourse & Society*, 9(4): 457–479.

Fairclough, N. 1993. "Critical discourse analysis and the marketization of public discourse: the universities". *Discourse & Society* 4(2): 133–168.

Han, Z. 2014. "The marketization of public discourse: The Chinese universities". *Discourse & Communication*, 8(1): 85–103.

Osman, H. 2008. "Re-branding academic institutions with corporate advertising: A genre perspective". *Discourse & Communication*, 2(1): 57–77.

Xiong, T. 2012. "Discourse and marketization of higher education in China: The genre of advertisements for academic posts". *Discourse & Society*, 23(3): 318–337.

Zhang, Y. and O'Halloran, K. L. 2013. "'Toward a global knowledge enterprise': university websites as portals to the ongoing marketization of higher education". *Critical Discourse Studies*, 10(4): 468–485.

The UK 'at risk': Social change and the proliferation of 'at risk' in print news media (1960-2015)

Jens O. Zinn
Lancaster University & University of Melbourne
j.zinn@lancaster.ac.uk, jzinn@unimelb.edu.au

The proliferation of *risk* words[22] has become a defining feature of public debates and news coverage in the UK after WW2. Linguists and sociologists became interested in examining the meaning of *risk* (Fillmore & Atkins 1992; Luhmann 1993) while some scholars have moved *risk* into the centre of the conceptualisation of modern society (Beck 1992, 2009). Even though a growing body of scholarship in sociology and linguistics have concentrated on risk issues, there has been almost no attempt to systematically examine historical changes of the discourse semantic of risk and their social roots (Zinn & McDonald 2018). This is surprising since both disciplines have emphasised that language and the social are inseparably connected (e.g. Berger & Luckmann 1969; Fairclough 1992; Halliday 1978).

This presentation reports from a diachronic study that builds on earlier work on MD-CADS (Partington 2010) to examine how the social realm and language are connected. It uses the *at risk* construct ('at * risk' with one or two words between 'at' and 'risk', and the hyphenized 'at-risk') as an example, and examines the social contexts of the significant increase of the (relative) numbers of 'at risk' constructs from the 1960s onwards in news coverage of UK print media. The study rests on The Times corpus as processed and made available by Andrew Hardie of CASS through Lancaster University's CQPweb server. It is complemented by a risk corpus composed of all articles containing a risk word of a selection of UK newspapers which provide digitised archives from 1960s or later until 2015 such as The Times (1960), The Financial Times (1982), the Guardian (1984), The Scotsman (1993), the Daily Mirror (1995) and the Daily Mail (1992). This risk corpus has also been processes and made available through the

---

[22] Risk words are defined as any lexical item whose root is risk (risking, risky, riskers, etc.) or any adjective or adverb containing this root (e.g. at-risk, risk-laden, no-risk; Zinn & McDonald, 2018: 70)

CQPweb server.

The analysis shows that in contrast to claims that recent environmental and technological risks were responsible for the proliferation of risk (Beck 1992), several developments in a range of different social domains are connected to the proliferation of 'at risk' in print news media in the UK such as economic (job security), political (industrial conflict), scientific (epidemiology), and environmental (flooding) changes. The presentation will outline these and other results in more detail such as, for example: (1) With the end of the *golden age* of welfare capitalism job insecurity became a common and continuing experience reported in the news. (2) With the introduction of the NHS the costs and service quality of the comprehensive health system has been the source of ongoing conflicts. However, the health sector is not only shaped by a negative prosody but new treatments for *at risk* patients. (3) However, it is the overall scandal that people put themselves unreasonably at risk or, even worse, that people are put at risk by others, which is newsworthy and is to a large degree responsible for the proliferation of the *at risk* compound across all domains from the 1960s onwards.

The presentation concludes finally suggests to complement available sociological theories to risk by an approach which interprets the proliferation of 'risk' as a discourse-semantic phenomenon which is underpinned by a range of different socio-linguistic developments.

References

Beck U (1992) *Risk Society: Towards a New Modernity.* London: Sage.
Beck, U (2009) *World at Risk*. Cambridge, UK; Malden, MA: Polity.
Berger PL & Luckmann T 1969 The Social Construction of Reality. A Treatise in the Sociology of Knowledge. Penguin.
Faiclough N 1992 *Discourse and Social Change.* Polity
Halliday, MAK 1978 Language as Social Semiotic. London: Edward Arnold.
Fillmore, C J and Attkins, B T (1992) Towards a frame-based lexicon: The semantics of RISK and its neighbors. In
Lehrer, A. & Kittay, E. (Eds.), *Frames, fields & contrasts: New essays in semantic & lexicon organization* (pp. 75-102). Hillsdale (NJ): Lawrence Earlbaum.
Hamilton C, Adolphs S and Nerlich B (2007) The meanings of 'risk': a view from corpus linguistics. *Discourse & Society,* 18: 163-81.
Hardie A (2012) CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3): 380-409.
Luhmann N (1993) *Risk: A sociological theory.* New York: A. de Gruyter.
Partington A 2010 'Modern Diachronic Corpus-Assisted Discourse Studies (MD-CADS) on UK newspapers: an overview of the project' *Corpora 5* (2): 83–108.
Zinn, J O (2010) Risk as Discourse: Interdisciplinary Perspectives. *CADAAD journal* 4(2): 106-124.
Zinn, J O and McDonald D (2018) Risk in The New York Times (1987-2014): A corpus-based exploration of sociological theories. Houndsmills, Basingstoke: Palgrave.

The medicalization of transgender identities in the British press

Angela Zottola
University of Napoli Federico II
zottolaangela@gmail.com

A study conducted on the representation of transgender identities in the British press in the time span that stretches between 2013 and 2015 (Zottola 2017) highlighted a set of discursive practices carried out by the press in the representation of transgender people. One of these

patterns pointed out to the fact that the newspapers in some cases had a tendency of associating transgender identity to a medical issue through the use of terms such as *dysphoria*, *sex change* and *gender reassignment surgery*. This correlation between trans identity and a medical condition was not always straightforward or explicit but ran between the lines, in the choice of specific terminology or contexts in which the topic was discussed; for example, the choice of the journalists to use *sex change* instead of *transition*. Considering the power of the press in influencing our understanding of given topics, these linguistics choices, if reiterated, could become a negative example and most of all a misleading practice for people who are not aware of the meaning of transgender identity and only come across this definition through the words of the newspapers they read.

In the framework of Corpus-assisted Discourse Studies (Partington 2004, 2012, 2015; Partington, Duguid & Taylor 2013) and drawing from Critical Discourse Analysis (Baker 2006; Baker Gabrieltaos and McEnery 2013 Bucholtz 2003; Fairclough 1995, 2005; van Leeuwen 2005) the aim of this contribution is to investigate a corpus of newspaper articles in order to critically address instances of medicalization of transgender identities in the press. In particular, the analysis will focus on the semantic prosody (Hunston 2007; Louw 1993; Partington 1998; Partington, Duguid and Taylor 2013; Sinclair 1991; Stubbs 1996) carried by the linguistic strategies preferred by the press when discussing trans identities. Moreover, the analysis will try to uncover the means through which the process of medicalization of this identity is enacted. The corpus under scrutiny, the TransCor (Zottola 2017), of more than 2 million words, was collected between January 2013 and December 2015. The articles included in the corpus were retrieved from eight national British newspapers: *The Guardian*, the *i*, *The Daily Telegraph* and *The Times* as representative of the quality press (Jucker 1992), and the *Daily Express*, the *Daily Mail*, the *Daily Mirror* and *The Sun* as representative of the popular press (Jucker 1992). In line with the choices of newspapers included in the corpus, this investigation will also address the differences in the representation given by the popular and quality press.

Preliminary findings highlight that despite a stronger presence of this medical related discourse is found in the popular press, this discursive practice is retraceable throughout the TransCor. Additionally, it was noted that the discourse not only varies across newspapers but also diachronically, in the small time span investigating. This last result demonstrates the increasing presence and relevance, therefore growing popularization, of this topic in the press.

References

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

Baker, P., Gabrielatos, C. and McEnery T., 2013. *Discourse Analysis and Media Attitudes. The Representation of Islam in the British Press*. New York: Cambridge.

Bucholtz, M. (2003). "Theories of Discourse as Theories of Gender: Discourse Analysis in Language and Gender Studies". In J. Holmes and M. Meyerhoff (eds) *The Handbook of Language and Gender* (43–68). Oxford: Blackwell.

Fairclough, N. 1995. *Critical Discourse Analysis. The Critical Study of Language*. London/New York: Longman.

Fairclough, N. 2005. *Analysing Discourse*. London/New York: Routledge.

Hunston, S. 2007. Semantic Prosody Revisited. *International Journal of Corpus Linguistics*, *12*(2), 249–268.

Jucker, A. 1992. *Social Stylistics*. Berlin/New York: Mouton de Gruyter.

Louw, B. 1993. "Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies". In M. Baker, G. Francis, and E. Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair* (157–176). Amsterdam: John Benjamins.

Partington, A. 1998. *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.

Partington, Alan S. 2004. "Utterly content in each other's company". Semantic prosody and semantic preference. *International Journal of Corpus Linguistics*, 9(1): 131–156.

Partington, Alan S. 2012. *The changing discourses on antisemitism in the UK press from 1993 to 2009: A modern-diachronic corpus-assisted discourse study. Journal of Language and Politics*, 11(1): 51–76.

Partington, Alan S. 2015. *Corpus-assisted Comparative Case Studies of Representations of the 'Arab world.* In *Corpora and Discourse Studies: Integrating Discourse and Corpora. London,* edited by P. Baker and T. McEnery, 220–243. Palgrave Macmillan.

Partington, A., Duguid, A., and Taylor, C. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-assisted Discourse Studies*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Sinclair, J. M. 1991. "Words and phrases". In J. M. Sinclair (ed.) Corpus, Concordance, Collocation (70–75). Oxford: Oxford University Press.

Stubbs, M. 1996. *Text and Corpus Analysis*. Oxford: Blackwell.

van Leeuwen, T. 2005. "Three models of interdisciplinarity". In R. Wodak, and P. Chilton (eds) *A New Agenda in (Critical) Discourse Analysis* (3–18). Amsterdam/Philadelphia: John Benjamins.

Zottola, A. 2017. *Transgender in the Spotlight. The Other-representation of Transgender People in the British Press: A Corpus-based Discourse Analysis.* Doctoral Thesis. University of Napoli Federico II.