

Book of Abstracts

7th Corpora & Discourse
International Conference



Organising Committee:

ANNA ISLENTYEVA
Department of English

KAROLINE IRSCHARA
Department of Linguistics

CHRISTINE KONECNY
Department of Romance Languages

CLAUDIA POSCH
Department of Linguistics

GERHARD RAMPL
Department of Linguistics

ELISABETH ZIMMERMANN
Department of English

Compiling:

LAURIN KLOCKER
JONAS OBERPARLEITER

Cover page and logo:

TATIANA ZHUKOVA

Layout:

MAHMUT BOZKURT
LEONIE GRAUPNER
FINN JAKOBS
ANNA NORZ
REKA PIHES
MARIE PLATTNER
HANNAH RIPPER
ALINA WALSER

Special thanks to our sponsors and supporters

UNIVERSITÄT INNSBRUCK

FACULTY OF LANGUAGE, LITERATURE AND CULTURE

INTERNATIONAL RELATIONS OFFICE

OFFICE OF THE VICE-RECTOR FOR RESEARCH

RESEARCH AREA "CULTURAL ENCOUNTERS – CULTURAL CONFLICTS"

RESEARCH PLATFORM "CENTER INTERDISCIPLINARY GENDER STUDIES INNSBRUCK"

RENATO SOFTWARE LTD.

LAND TIROL

LAND VORARLBERG

Index of Authors

Keynote

Baker, Paul 5
Mautner, Gerlinde 6

Stefanowitsch, Anatol 7
Vessey, Rachelle 8

Papers (primary authors)

Alcaraz-Mármol, Gema 10
Archer, Dawn 12

Beauchamp, David 14
Bevitori, Cinzia 16

Biri, Ylva 18
Bisiada, Mario 20

Borza, Natalia 22
Brocca, Nicola 24, 26
Bruce, Linda 28

Cancio Pastor, Tessa 30
Cantos Delgado, Clara 32

Clarke, Ben 34
Clarke, Isobelle 36

Conti, Sergio 37
Currie, John 39

Curry, Niall 41

Dayter, Daria 43
de Lima Lopes, Rodrigo Esteves 44

Del Fante, Dario 46
Diegoli, Eugenia 48

Diemer, Stefan 50
Duguid, Alison 52

Dumlao, Remart Padua 54
Dykes, Nathan 55

Eyssette, Sophie 67

Fábián, Annamária 60
Falcone, Mariasophia 62
Feldmüller, Tim 64

Gillings, Mathew 66
Giordano, Luca 68
Güler, Aslıhan Tuğçe 73

Hathaway, Yulia 75
Hoffmann, Christian 77
Huang, Jing 79

Imamovic, Mirela 81
Iori, Ilaria 83

Kania, Ursula 85
Kaukonen, Elisabeth 87
Kennedy, Charlotte-Rose 89

Kirner-Ludwig, Monika 91
Klocker, Laurin 92

Kopf, Susanne 94
Kurkina, Arina 96

Leone, Ljubica 98
Lischinsky, Alon 100

Liu, Siyuan 103
Locatelli, Laura 105
Lustig, Andrew 107

Marchi, Anna 109, 111
Marko, Georg 113

Mateja, Dagmara 117
Mayr, Paul Valentin 119

Meyer, Anna 121
Mukherjee, Sarah Jane 123

Narisano, Laura 125
Ngula, Richmond Sadick 127
Nikitina, Jekaterina 129

Orrequia-Barea, Aroa 131

Parnell, Tamsin 133
Pelizzari, Nicola 135

Poole, Robert 137
Posch, Claudia 138
Prado-Alonso, Carlos 140

Rüdiger, Sofia 142
Ryoo, Mi-Lim 144

Sánchez Cócera, Marta 146

Taylor, Charlotte 148
Trklja, Aleksandar 150
Tubaltseva, Svitlana 152

Verachtert, Freja 154
Villegas, Felipe 156

Wong, Catherine 159

Zaupa, Federico 161
Zhao, Xiaowen 163

Zhu, Yafei 165
Zimmermann, Elisabeth Katharina 167
Zipser, Katharina 169

Posters (primary authors)

Biber, Hanno 171

Bogetic, Ksenija 172

Combei, Claudia Roberta 174

D'Agostino, Dario 175

Disdier, Johanna 177

Elemerot, Irene 179

Fitzpatrick, Iona 181

Hallin, Katherine 183

Kania, Ursula 185

Kiemes, Carina 187

Landwehr, Isabell 189

Leedham, Maria 191

Lukač, Morana 193

Marko, Karoline 194

Pfeifer, Ulrike 195

Risberg, Lydia 197

Sano, Akira 199

Suremann, Isabelle 203

Szutta, Zuzanna Maria 205

Added value or perilous activity?

The affordances and challenges of automatic annotation for CADS

Paul Baker (Lancaster University, United Kingdom)

Automatic annotation is the process of using computer programs to assign codes or tags to a corpus, based on a pre-existing categorisation scheme, supplementing texts with additional levels of information. While a popular technique in areas of corpus linguistics like register variation and grammatical analysis, it is less commonly used in CADS-based research, where analysts have tended to either use unannotated corpora or hand-tag their corpora, employing their own categorisation schemes based on specific research goals. In this talk I describe a range of tagging systems and accompanying software and then provide case studies from my own research which have used various automatic taggers, reflecting on what was gained and lost from taking this approach. I also discuss more recent developments in automatic tagging, including the image tagger Google Cloud Vision and the use of AI tools like ChatGPT to implement bespoke tagging systems. I conclude that automatic tagging can provide useful insights for CADS analysts while outlining a number of important caveats to consider.

The marriage of true minds ... and some impediments

Reflections on CADS in theory and practice

Gerlinde Mautner (Wirtschaftsuniversität Wien, Austria)

CADS has arrived. For many discourse analysts, it is now the go-to method for studying large, representative datasets rather than small samples of purposefully selected texts. And indeed, Computer-Assisted Discourse Studies promises to deliver the best of both worlds: quantitative breadth and qualitative depth.

That said, significant methodological challenges remain, such as the inherent bias of the approach towards lexical and local phenomena; the focus on verbal data at the expense of multimodality; the tendency for the quantitative part of the analysis to encroach upon the qualitative component; and over-confidence in computer-generated 'findings' without careful contextualisation and interpretation.

While I will argue in my talk that none of these pitfalls and limitations detract from CADS' great potential, I will also make a strong case for embracing a discerning, cautious and critical attitude towards it. In some cases, research designs really do benefit immensely if a CADS element is added; in others, research designs are better off without it. The challenge (obviously) is to know the difference.

If we look beyond linguistics, however, we find that many researchers working with discourse data do not even consider using CADS, so that the question of weighing up its advantages and drawbacks does not even arise. In management studies, for example, the approaches used most frequently are content analysis on the one hand and topic modelling on the other. Corpus linguistics and CADS are often simply not on researchers' radar screen.

Together with colleagues from the UK, I've recently tried to reach out to these scholarly communities, attempting to convince them that linguistics generally and CADS in particular would make a useful addition to their conceptual and methodological armoury. Reporting on this work, I will conclude by outlining an interdisciplinary research agenda that aims to establish CADS more firmly in the methodological canon of the social sciences.

Populism as politics, narrative and trope

A call for an integrated discourse- and corpus-linguistic approach to populism

Anatol Stefanowitsch (Freie Universität Berlin, Germany)

In discourse-analytic approaches to political populism, at least three phenomena need to be distinguished: first, populist policies and politics, second, populist framings and third, populist linguistic tropes (expressions instantiating these narratives). While we may assume a correlation between the three phenomena, they do not necessarily co-occur: a party can (claim to) pursue populist policies without using populist narratives, or it may use populist narratives without pursuing or proposing any populist policies and populist narratives may or may not manifest as populist linguistic tropes. This makes the study of populist discourses more complex than many scholars have so far assumed — specifically, using a single method of analysis is likely to yield an incomplete or distorted picture. I will illustrate this problem based on a qualitative and quantitative analyses of five years of press releases by the German Alternative für Deutschland. This far-right party is usually described as “populist” by political scientists (and by some discourse analysts); however, a closer look reveals a far more complex situation: many (perhaps most) of the party’s policies are not straightforwardly populist and where they are populist, they are mostly presented without the use of standard populist tropes. However, by integrating qualitative discourse-analytic approaches and quantitative corpus-linguistic methods into a broader historical and political analysis we can uncover a range of more or less subtle populist narratives and tropes.

Intersectional CADS

Rachelle Vessey (Carleton University, Canada)

Intersectionality refers to a research perspective in which a *multiplicity* of identity-based positions and subjectivities is foregrounded over a focus on *singular* identity categories. While intersectional perspectives have been increasingly adopted in linguistic research (Levon 2015), there has been little adoption amongst CADS scholars (although see Jaworska and Hunt 2017; de la Ossa 2019).

Intersectionality scholar Kimberlé Crenshaw (1991: 1249) has highlighted that access to resources and support can be contingent on the adoption of a particular language; as such, *language use* is a potentially important intersectional consideration. In this paper, I argue for an intersectional CADS approach that starts by addressing (and distinguishing between) *linguistic* and *discursive* choices. By “linguistic choice”, I attend to the significance of multilingualism, which is often overlooked within corpus research (Nurmi and Rütten 2017). By “discursive choice”, I refer to the semiotic activity through which we represent our meanings (Halliday 2013). Neither linguistic nor discursive choices are neutral and both contribute to meaning making. This paper aims to highlight how the intersection of linguistic and discursive choices produces non-obvious meanings (Partington 2017).

I propose that an intersectional CADS approach can begin with the identification of meanings encoded in a corpus via the languages contained within it. Then, the analysis can proceed by identifying how language choices align with discursive constructions (and onward to other identity-related considerations). I offer two cases where I apply CADS in different ways to examine intersections of language choices and discursive choices, exploring implications for identity-based positionality and subjectivity. The first case addresses a small corpus of United Nations deliberations about multilingualism between 1995-2022 (56,563 words), which took place in the six official languages of that organization (McEntee-Atalianis and Vessey in progress). In this study, corpus annotation is used to access intersections of language choice and the discursively marked positionality of UN member states. The second case is drawn from CADS research on the language of sexual violence and impropriety in the Canadian Armed Forces (1.6 million words), in which both official languages of Canada are used (Vessey forthcoming). Here, corpus segmentation and inter- and intra-corpus comparisons are used to establish how sexual violence and impropriety are discursively constructed on the basis of gender, military rank and language.

References

- Candelas de la Ossa, Abigael. (2019). Exceptionalising intersectionality: A corpus study of implied readership in guidance for survivors of domestic abuse. *Gender and Language* 13(2): 224-250. [10.1558/genl.35094](https://doi.org/10.1558/genl.35094)
- Crenshaw, Kimberlé. (1991). Mapping the margins: Intersectionality, identity politics and violence against women of color. *Stanford Law Review* 43(6): 1241-1299. [10.2307/1229039](https://doi.org/10.2307/1229039)
- Halliday, Michael Alexander Kirkwood. (2013). Meaning as choice. In *Systemic Functional Linguistics: Exploring Choice*. Lise Fontaine, Tom Bartlett and Gerard O'Grady (eds.) Cambridge: Cambridge University Press, 15-36.
- Jaworska, Sylvia and Sally Hunt. (2017). Differentiations and intersections: A corpus-assisted discourse study of gender representations in the British press before, during and after the London Olympics 2012. *Gender and Language* 11(3): 336-364.
- Levon, Erez. (2015). Integrating intersectionality in language, gender and sexuality research. *Language and Linguistics Compass* 9(7): 295-308.
- McEntee-Atalianis, Lisa and Rachelle Vessey. (in progress). *Agency and positioning in United Nations language policy*.
- Nurmi, Arja and Tanja Rütten. (2017). How many languages are there in a monolingual corpus? In *Challenging the Myth of Monolingual Corpora*. Arja Nurmi, Tanja Rütten and Päivi Pahta (eds.) Leiden: Brill Publishing, 1-15.
- Partington, Alan. (2017). Varieties of non-obvious meaning in CL and CADS: From 'hindsight post-dictability' to sweet serendipity. *Corpora* 12(3): 339-367.
- Vessey, Rachelle. (forthcoming). *Language, sexual violence and impropriety in the military*. Cambridge: Cambridge University Press.
- Vessey, Rachelle. (2013). Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora* 8(1): 1-26. [10.3366/cor.2013.0032](https://doi.org/10.3366/cor.2013.0032)

Empowerment vs. victimization

Spanish-English semantic preferences in portraying Afghan women during the Taliban takeover

Gema Alcaraz-Mármol (Universidad de Castilla-La Mancha, Spain)

Jorge Soto-Almela (Universidad de Alicante, Spain)

Keywords: social media, corpus analysis, semantic preference, Afghanistan, woman

Social media, a pivotal source of information and personal expression, has been extensively examined in linguistic research over the past decade. The studies in this area often employ varied methodologies like (critical) discourse analysis, multimodal analysis and corpus analysis, as further elaborated by Sun, Wang and Feng (2021). Twitter – now X –, a platform for broadcasting news and opinions on current events, has garnered the attention of linguists (Anderson and Huntington 2017; Lutzky and Lawson 2019; Breeze 2020; Nugraha et al. 2021). The focus on women’s representation, whether in contexts such as sports (Adá Lameiras 2019), video games (Huntemann 2015), International Women’s Day (Gray 2019), women’s driving rights in Saudi Arabia (Altoaimy 2018), or elections with female candidates (Ruiloba-Nuñez and Goenaga 2022), highlights their diverse portrayal. Meanwhile, Twitter activism often presents women as empowered and challenging stereotypes, countering sexist discourse (Sampietro, Calvo and Campos 2020; Palomino-Manjón 2022; Williams and Gonlin 2017; Rentschler 2015).

This research utilized Twitter to create a corpus of tweets about a globally impactful event in August 2021: the withdrawal of international troops from Afghanistan and the subsequent Taliban interim government announcement. The Taliban’s takeover of Kabul on August 15, 2021, following President Ashraf Ghani’s flight to the United Arab Emirates, had significant repercussions, especially for women, whose rights were under threat.

The study aims to analyse and contrast the portrayal of Afghan women in tweets from Spain and the United States, focusing on semantic preference during the critical 17-day period from the Taliban’s Kabul takeover to the Afghan borders’ closure. Tweets in Spanish (ES Corpus) and English (US Corpus) containing the hashtags and keywords *Afganistán/Afghanistan* were collected through the Twitter APIv2, amounting to over 5,000 Spanish and 16,000 English tweets. Using WordSmith (Scott 2012), content words were extracted after applying a stop-list. The lemmas *mujer* and *woman* featured prominently, with a higher ranking in Spanish. Their semantic preference was analysed through collocates, as described by Bednarek (2008), involving semantic labelling and classification in each language.

The results seem to indicate that in both the ES and US Corpora, the lemmas *mujer* and *woman* are among the top 10 words in discussions on the Taliban’s Kabul takeover. This highlights the prominence of women in an apparently warlike discourse. The semantic preferences for *mujer* and *woman* span three categories: empowerment,

mistreatment and benevolence-empathy, with empowerment more prevalent in the ES Corpus (24.85%) and benevolence-empathy in the US Corpus (20.32%). This study reveals a dual narrative of victimization and empowerment in representing Afghan women on Twitter during the Taliban takeover of Kabul.

References

- Adá Lameiras, Alba. (2019). Invisibilización de La Mujer Deportista en el Twitter de los medios deportivos. *Comunicación Y Género* 2(1): 33-46. 10.5209/cgen.64460
- Altoaimy, Lama. (2018). Driving change on twitter: A corpus-assisted discourse analysis of the Twitter debates on the Saudi ban on women driving. *Social Sciences* 7(5): 81. 10.3390/socsci7050081
- Anderson, Ashley A. and Heidi E. Huntington. (2017). Social media, science and attack discourse: How Twitter discussions of climate change use sarcasm and incivility. *Science Communication* 39(5): 598-620. 10.1177/1075547017735113
- Bednarek, Monika. (2008). Semantic preference and semantic prosody re-examined. *Corpus Linguistics and Linguistic Theory* 4(2).10.1515/clt.2008.006
- Breeze, Ruth. (2020). Exploring populist styles of political discourse in Twitter. *World Englishes* 39(4): 550-567. 10.1111/weng.12496
- Gray, Daniel. (2019). *Review of tweeting about women: A critical discourse analysis of international women's day on Twitter*. PhD Thesis. Cardiff University.
- Huntemann, Nina. (2015). No more excuses: Using Twitter to challenge the symbolic annihilation of women in games. *Feminist Media Studies* 15(1): 164–167. 10.1080/14680777.2015.987432
- Lutzky, Ursula and Robert Lawson. (2019). Gender politics and discourses of #mansplaining, #manspreading and #maninterruption on Twitter. *Social Media + Society* 5(3): 1-12. 10.1177/2056305119861807
- Nugraha, Intan Siti, Eva Tuckyta S. Sujatna and Sutiono Mahdi. (2021). Corpus linguistic study of tweets using #charlieheβδο hashtag. *Journal of Applied Linguistics and Literacy* 5(1): 54–70. 10.25157/jall.v5i1.4965
- Palomino-Manjón, Patricia. (2022). Feminist activism on Twitter. The discursive construction of sexual violence and victim-survivors in #WhyIDidntReport. *Journal of Language Aggression and Conflict* 10 (1): 140-168. 10.1075/jlac.00049.pal
- Rentschler, Carrie. (2015). #Safetytipsforladies: Feminist Twitter takedowns of victim blaming. *Feminist Media Studies* 15(2): 353–56. 10.1080/14680777.2015.1008749
- Ruiloba-Núñez, Juana María and María Goenaga Ruiz de Zuazu. (2022). Género, estilos de liderazgo y campañas electorales en Twitter: Las elecciones Andaluzas de 2018. *Opinião Pública* 28(2): 510-532. 10.1590/1807-01912022282510
- Sampietro, Agnese, Dafne Calvo and Eva Campos. (2020). Los emojis del 8M: Su uso en Twitter durante las movilizaciones feministas de 2019. *Revista Dígitos* 1(6): 137–158. 10.7203/rd.v1i6.170
- Scott, Mike. 2012. WordSmith Tools [Version 6.0.] <http://www.lexically.net/wordsmith/downloads>
- Sun, Ya, Gongyuan Wang and Haiying Feng. (2021). Linguistic studies on social media: A bibliometric analysis. *SAGE Open* 11(3): 1–12.10.1177/21582440211047572
- Williams, Apryl and Vanessa Gonlin. (2017). I got all my sisters with me (on Black Twitter): Second screening of *How to Get Away with Murder* as a discourse on black womanhood. *Information, Communication & Society* 20(7): 984–1004. 10.1080/1369118x.2017.1303077

The mainstreaming of misogyny?

Andrew Tate's prevalence in corpora from Reddit and the British press

Dawn Archer (Manchester Metropolitan University, United Kingdom)

Frazier Heritage (Manchester Metropolitan University, United Kingdom)

Keywords: misogyny, mainstreaming, manosphere, semantic domain analysis, interdiscursivity

Journalists have recently highlighted growing concerns over the mainstreaming of misogynistic ideologies due to influencers like Andrew Tate (e.g. Kingsley 2023). To date, little linguistic research has been done to confirm – and, where relevant, explore the prevalence of – such mainstreaming (though see Solea & Sugiura 2023 for a discussion of the mainstreaming of male-supremacist views more broadly). This paper thus explores Tate and his ideologies by investigating the degree to which he/they are reflected in corpora from Reddit and the British press.

We initially explored intertextual references to Tate in a 7.7-million-word corpus of Reddit data scraped in October/November 2023. We found that explicit reference to him, his actions and/or ideologies only occurred 9 times. Tate and his ideologies were appraised negatively in every occurrence, in relation to women, veracity, likeability, masculinity and/or money.

This led us to explore intertextual references to Tate in a 1,004,420-word corpus, representative of 2023, generated from Lexis Nexis. We found that he appeared 6,358 times across 1211 articles, with the press typically framing him in negative ways: they exhibited particular concern regarding the degree to which they perceived his ideologies to be being mainstreamed and the influence this might have on young people.

As our preliminary findings indicated that intertextual references typically challenged Tate's ideologies/appraised him negatively, we also explored whether his ideologies were being mainstreamed through interdiscursive references. We scraped a sample of 6,000 words from Tate's professional website to, first, determine his use of self-presentation and other presentation, under two assumptions: that such techniques are part of impression management (henceforth IM) and that IM is a well-known influence/manipulation strategy, given that it equates to the 'strategic management of behaviour (and, potentially, appearance) in order to influence the perceptions of others' (Goffman 1959), (Archer 2018: 205). A key semantic field analysis (Rayson 2008) reveals that Tate typically discusses (views of) women, (mentions of) veracity/credibility, (words associated with) likeability, (hyper) masculinity and (the importance of) money: that is, the things he is often criticised for (see above). However, initial comparisons of the differing corpora suggest that these semantic domains are constructed in very different ways by Tate, with him constructing them especially positively (thereby providing us with a way into his different worldview to the "mainstream"). We conclude, from this, that Tate's views and thus influence, are not as "mainstream" as the British press appear to fear.

Based upon the findings, we highlight two needs. The first relates to the importance of critically reflecting upon any presuppositions relating to an influencer's "reach", "significance", degree to which they are (becoming) "mainstream", etc. The second arises out of other available reference corpora of Reddit data being more than 5 years old. We believe there is a need for such reference corpora to be frequently updated – as well as being made available in some capacity –so that it is possible to study whether people like Tate become "newsworthy" for a "season" only or over time.

References

- Archer, Dawn. (2018). Impression management in the early modern English courtroom. In Dialogues in Diachrony. Merja Kytö and Terry Walker (eds.). *Journal of Historical Pragmatics* 19(2): 205-222.
- Goffman, Erving. (1959). *The presentation of self in everyday life*. Garden City, New York: Anchor/Doubleday.
- Kingsley, Thomas. (2023). Experts warn of 'copycat Andrew Tates' as influence continues to grow from behind bars. The Independent. <https://www.independent.co.uk/news/uk/home-news/andrew-tate-arrest-romaniantrafficking-influencer-b2275233.html> (Accessed January 2024)
- Rayson, Paul. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics* 13(4): 519-549.
- Solea, Anda Iulia and Lisa Sugiura. (2023). Mainstreaming the blackpill: Understanding the incel community on TikTok. *European Journal on Criminal Policy and Research* 29(3): 311-336.

Linguistic variation in NHS tweets throughout the first year of the pandemic

National and regional findings

David Beauchamp (Coventry University, United Kingdom)

Keywords: coronavirus, health communications, crisis communications, sentiment analysis, social media

The sudden onset of the coronavirus pandemic necessitated rapid and widespread communication from institutions to the public. To this end, the use of social media by health organisations was extensive. Although there has been an abundance of research into coronavirus-related social media communication in the first year of the pandemic, most of this research focused upon the initial short-term period of the first few weeks and months of 2020. There has been little research into long-term strategies of social media-based public health messaging during the coronavirus crisis. Of particular interest is the extent to which it may have aligned with or departed from existing frameworks of best practice for health communications, which recommend health communications be engaging, credible, relevant, informative and dialogic (Morgan 2009; Schiavo 2013; Malecki 2020).

This research paper examines the linguistic content of the tweets of two NHS twitter accounts during the coronavirus pandemic: one the national level twitter account of the NHS (@NHSuk) and the other a regional level account based in Northwest England (@NHSNW). Its assumption of variation in the nature of the language used is rooted in the Hallidayan/Functional approach to language (Halliday & Matthiessen 2014). This posits that the changes in social situation over time and differences between the communicative goals of national and regional NHS representatives, will be realised through variation in linguistic output.

A corpus of 2563 tweets were collected from March 2020 to March 2021. Corpus linguistic software (Sketch Engine) and R are used to analyse the tweets in terms of their keyness, sentiment polarity and complexity and how these changed throughout the first year of the pandemic. Trends identified through quantitative and statistical approaches are then examined more closely at the level of the individual tweet for a more in depth understanding of communicative practices and how these evolved over time.

The results reveal some striking similarities and differences between the linguistic output of the two twitter accounts. Both sets of tweets show similar use of positive sentiment during times of crisis and more negative sentiment at times of relief. However, there are also clear differences in strategies of engagement and key themes, with the national account focusing more on generic form and more scientific content, while the regional account shows greater variation in form and more emphasis on engagement with relevant stakeholders and other members of the public.

It is hoped that the findings of this research can inform best practice for institutional communications teams to devise effective strategies which can be deployed in the event of future, similar crises.

References

Halliday, Michael Alexander Kirkwood and Christian M. Matthiessen. (2014). *An introduction to functional grammar*. Routledge.

Malecki, Kristen, Keating, Julia and Nasia Safdar. (2020). Crisis communication and public perception of COVID-19 risk in the era of social media. *Clinical Infectious Diseases* 72(4): 699–704. 10.1093/cid/ciaa758.

Morgan, Michael. (2009). Cultivation analysis and media effects. In *The Sage handbook of media processes and effects*. Robin L. Nabi and Mary Beth Oliver (eds.) Thousand Oaks, CA: Sage, 69–82.

Schiavo, Renata. (2013). *Health communication: From theory to practice*. 2nd ed. Jossey-Bass.

What's in a name?

Representing aphasia and autism over time: A CADS perspective

Cinzia Bevitori (Università di Bologna, Italy)

Guilia Bencini (Università Ca' Foscari Venezia, Italy)

Keywords: aphasia, autism, referential strategies, corpus-assisted critical discourse study

There is a current debate in disability discourse with respect to the socio-cultural meanings of two alternative linguistic referential strategies in English: adjectival premodification (*disabled/blind/deaf/autistic* person) or post-nominal modification with a prepositional modifier and a nominal element that specifies the nature of the disability (*person with a disability/low vision/hearing impairment/autism*). These referential options are known as identity-first and person-first language. The rationale for person-first language is to focus on personhood as the central notion, therefore not identifying a person with their disability. Proponents of identity-first language, instead, claim that pre-modifying syntax affirms disability as essential to one's lived experience and identity. Previous corpus-based studies have examined identity-first vs. person-first language as a window onto how disability is represented in newspaper discourse, either examining disability in general (Potts et al. 2023), or focusing on one specific condition (Karaminis et al. 2023)

We present the results of a diachronic corpus-assisted study comparing two health conditions: aphasia and autism, under the hypothesis that the shift towards identity-first is not uniform across conditions and depends on the existence of recognized communities. Aphasia is an acquired language and communication disorder and is typically conceptualized as an impairment. Autism is a congenital neurodevelopmental condition, increasingly seen as a difference, not an impairment. Unlike aphasia, autism has a growing identity affirming community within the broader neurodiversity movement. Our research questions are as follows: Are there similarities and/or differences in the use of person-first vs. identity first language for autism vs. aphasia? Are there evolutions over time? We expected to find differences in the use of referential expressions and their discourse/evaluative prosodies. We used a purpose-built diachronic corpus (January 2000 – December 2022). We selected two comparable scientific journals: *Aphasiology* and *Autism* and randomly sampled two issues per journal per year. The corpus consists of 568 texts and approximately 4,107,000 running words. We performed quantitative and qualitative analyses using Sketch Engine and Wordsmith Tools 8. We computed the frequency and distribution of select search terms. For 'aphasia', these were: *with *aphasia*, *PWA* (acronym of *person/people with aphasia*), the noun *aphasic/s* and the adjective *aphasic*. Search terms for 'autism' were: *with *autism*, *ASD* (acronym for *autism spectrum disorder*) and the noun *autistic/s* and the adjective *autistic*. We used word sketches and concordances to identify collocational profiles for each term.

As predicted, quantitative analyses of the frequency distributions of identity-first and person-first language revealed differences over time and across conditions. On average, we observed an increase of person-first language for aphasia (i.e., from aphasic to person-with-aphasia). For autism, the shift was in the opposite direction, with an overall increase in the use of identity-first language (i.e., from person with autism, to autistic). A qualitative analysis of their collocational profile and patterns in their lexico-grammatical contexts showed a meaningful shift in their discourse prosodies over time. We will further discuss the implications of these findings with respect to evolving disability frameworks (Shakespeare & Watson 2010; Grue 2015), within and across different communities of practice.

References

- Grue, John. (2015). *Disability and discourse analysis*. London: Routledge.
- Karaminis, Themis, Costas Gabrielatos, Ursula Maden-Weinberger and Geoffrey Beattie. (2023). Portrayals of autism in the British press: A corpus-based study. *Autism* 27(4): 1092-1144.
- Potts, Amanda, Bednarek, Monica and Annemarie Watharow. (2023). Super, social medical: Person-first and identity-first representations of disabled people in Australian newspapers, 2000–2019. *Discourse and Society* 34(4).
- Shakespeare, Tom and Nicholas Watson. (2010). Beyond models: Understanding the complexity of disabled people's lives. In *New directions in the sociology of chronic and disabling conditions: Assaults on the lifeworld*. Graham Scambler and Sasha Scambler (eds.) London: Palgrave Macmillan, 57-76.

Professionalism and positioning in investment advice on YouTube

Ylva Biri (University of Helsinki, Finland)

Keywords: self-presentation, positioning, YouTube, influencers, instructional discourses

This paper discusses positioning and interdiscursivity by YouTube influencers offering financial advice. These so-called ‘finfluencers’ are a hitherto understudied group of content creators, often lacking institutionally sanctioned credentials, whose output is consulted by young people in particular and may thus impact their financial literacy and understanding of wealth and financial success (Guan 2023). Besides acting as experts in their profession, influencers seek to be engaging and authentic (e.g., Chan 2019; Tolson 2010). Their self-presentation can be seen as situational and discursively constructed (Davies & Harré 1990) and as interdiscursive in so far as it addresses the different needs and practices of the communicative context of a YouTube video (e.g. Bhatia 2018). This paper asks (1) what kind of situational discursively constructed performances or ‘positions’ YouTube finfluencers construct discursively and (2) what macro-level (discursive) self-performances or discourses these positions contribute to over the course of a video.

The corpus data of the analysis contains 123,000 tokens from 23 English-language YouTube videos, representing 8 hours of footage. The videos represent tutorials provided by the YouTube algorithm to the search query “how to invest”. The strategies of self-presentation selected for the paper are identified through corpus-assisted discourse analysis. For identifying situational positioning, a corpus query of the pronouns “I”, “we” and “you” is used to identify identity-labels and -claims of the influencer in relation to their audience. This quantitative approach operationalises positioning strategies as lexico-grammatical constructions while allowing the statistical comparison of positions. For a more contextualised view of the positions, segments in the instructional videos are manually analysed in terms of their rhetorical goal.

While a YouTube video contributes to a more stable professional “self” of the finfluencer, a single video contains multiple situational positions connecting to varying discourse goals. The functions of linguistic positioning strategies in the corpus vary in terms of how pronouns are used for self- and otherpositioning to involve and engage their audience (Biri 2021). Interdiscursive awareness (Bhatia 2018) is evident as the finfluencers build rapport with their audiences for example by self-positioning as approachable and relatable. There are also differences in what rhetorical goals or practices their videos include, but generally a video switches between explaining, demonstrating, or narrating the influencer’s own experiences. In the finfluencer context, many positions also relate to narratives of the influencer’s own experiences or the viewer’s hypothetical experiences or expectations. These segments contribute to ideological discourses by presupposing that striving towards wealth is desirable, as the influencer positions themselves as a role-model of someone who has attained wealth through hard work. The findings contribute to the growing research on YouTuber discourse while pioneering the new topic of “finfluencers”,

whose discourses on consumption and lifestyle differ from that of, e.g., beauty influencers, a researched but more female audience-oriented discourse community.

References

- Bhatia, Aditi. (2018). Interdiscursive performance in digital professions: The case of YouTube tutorials. *Journal of Pragmatics* 124: 106-120. 10.1016/j.pragma.2017.11.001
- Biri, Ylva. (2021). Metadiscourse in social media: A reflexive framework. In *Metadiscourse in digital communication: New research, approaches and methodologies*. Larissa D'Angelo, Anna Mauranen and Stefania Maci (eds.) Springer International Publishing, 133-54.
- Chan, Ngai Keung. (2019). Becoming an expert in driving for Uber: Uber driver/bloggers' performance of expertise and self-presentation on YouTube. *New Media & Society* 21(9): 2048-67. 10.1177/1461444819837736
- Davies, Bronwyn and Rom Harré. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behaviour* 20(1): 43-63. 10.1111/j.14685914.1990.tb00174.x
- Guan, Sue S. (2023). The rise of the finfluencer. *New York University Journal of Law & Business* 19(3): 489.563. 10.2139/ssrn.4400042
- Tolson andrew. (2010). A new authenticity? Communicative practices on YouTube. *Critical Discourse Studies* 7(4): 277-89. 10.1080/17405904.2010.511834

Migration and metaphors of water

A contrastive corpus-assisted study

Mario Bisiada (Universitat Pompeu Fabra, Spain)

Dario Del Fante (Università degli studi di Ferrara, Italy)

Irene Elmerot (Stockholms Universitet, Sweden)

Charlotte Taylor (University of Sussex, United Kingdom)

Keywords: metaphors, collocations, cross-linguistic CADS, migration discourse

Existing research shows that metaphors from water domains are frequently and cross-linguistically used to talk about migration phenomena (Charteris-Black 2006; Tkaczyk et al. 2015; Porto 2022; Taylor 2020, 2022). Much of this work has been conducted on countries that have sea access, so that talking about “waves of immigrants” may be part of those countries’ embodied connection to seafaring. This has given rise to questions about the generalisability of these findings and so there is still work to do on landlocked countries. Such countries, in theory, do not have a marine cultural heritage (and embodied experience) to the same extent and might be expected to have a different discursive conceptualisation of water metaphors. This talk examines issues around water metaphors and the embodiment hypothesis by asking: does living in a country with sea access influence the use of water metaphors? A wider question addressed by this research is to what extent discursive strategies are used like Foucault’s (1972) epistemes across languages, irrespective of their conceptual mappings (e.g. national discourse).

We focus on water metaphors in political debates from four countries, two landlocked and two coastal: Austria, Czechia, Italy and the United Kingdom. To analyse these four countries, we use the multilingual comparable ParlaMint corpora (Erjavec et al. 2023), containing transcriptions of parliamentary debates 2015–2022, a period which in Europe involves two intense immigration phases.

The analysis combines corpus linguistic and discourse analysis (Baker 2023). Firstly, we identify cooccurrences of names for people who move (e.g. *refugees*, *asylum seekers*, *immigrants*, *migrants* and functional equivalents) and lexical items which might refer to water (e.g. *flow*, *wave* and equivalents). We undertook this independently to ensure that we had a ‘bottom up’ approach to the metaphors in each language corpus. Secondly, we analyse these collocational patterns and concordances in each of the four corpora to identify the targets of water metaphors (i.e. who is being described as water), the evaluation offered and the discourses invoked (e.g. natural resource or natural disaster). Lastly, we compare the findings across the four contexts, drawing out similarities and differences and interpreting these findings to ask whether the discursive and conceptual proximity of migration and water – people arriving by the sea – influence the metaphorical production. Tentative results show that Austrian has three popular water metaphors, two of which are based on riverine terminology, Czech has four popular ones, of which three are riverine, English presents eight

of which four are riverine and Italian has four popular water metaphors, of which two based on riverine terminology. In general, there are no water metaphors exclusively occurring with just one topic, but for Austria, words meaning “stream” have mainly migrational collocations. For Czech, both “flood” and “wave” have mainly migrational collocations in the whole corpus, whereas the noun and verb for “stream” also have other topics. For English, the nouns “torrent” and “deluge” have other collocational topics than migration, whereas “influx, tide, stream, surge, flood, wave” have migrational collocations. For Italian, the nouns “wave”, “flow” and “influx” mainly occur within migration discourse.

References

- Baker, Paul. (2023). *Using corpora in discourse analysis*. 2nd ed. London: Bloomsbury.
- Charteris-Black, Jonathan. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society* 17(5): 563-581. 10.1177/0957926506066345
- Erjavec, Tomaz et al. (2023). Linguistically annotated multilingual comparable corpora of parliamentary debates ParlaMint.ana 4.0. <http://hdl.handle.net/11356/1860>
- Foucault, Michel. (1972). *The archaeology of knowledge*. New York: Pantheon Books.
- Porto, Dolores M. (2022). Water metaphors and evaluation of Syrian migration: The flow of refugees in the Spanish press. *Metaphor and Symbol* 37(3): 252–267. 10.1080/10926488.2021.1973871
- Taylor, Charlotte. (2020). Representing the Windrush generation: Metaphor in discourses then and now. *Critical Discourse Studies* 17(1): 1–21. 10.1080/17405904.2018.1554535
- Taylor, Charlotte. (2021). Metaphors of migration over time. *Discourse & Society* 32(4): 463-481. 10.1177/0957926521992156
- Taylor, Charlotte. (2022). The affordances of metaphor for diachronic corpora & discourse analysis: Water metaphors and migration. *International Journal of Corpus Linguistics*. 10.1075/ijcl.22004.tay
- Tkaczyk, Michael, Pospěch, Pavel and Jakub Macek. (2015). *Analýza mediálního pokrytí uprchlické krize (výzkumná zpráva)*. Masarykova univerzita. 10.13140/RG.2.1.4957.5764

We won't let the heritage of many thousand years be lost!

Discursive legitimization strategies in the Facebook comments of Fidesz supporters

Natalia Borza (Pázmány Péter Catholic University, Budapest, Hungary)

Keywords: discursive legitimization, social media, comment, values and norms, populism

According to a widely accepted definition of populism (Mudde & Kaltwasser 2017: 6), it is an ideology which “argues that politics should be an expression of the *volonté générale* (general will) of the people”. This paper investigates the motivations of the supporters of a right-wing populist party in Hungary. The present study focuses on the discourse of the addressee of a populist party. The research uncovers what motivated people to vote for the governing party *Fidesz* to act as a representative of the Hungarian people in an international context in the European Parliament elections in 2019. The study seeks answers to the research question: What values played a role in shaping the voting preference of the party's supporters?

In order to map the motivations of the supporters, the following methodology was applied on a specifically built corpus. Postings that campaigned for the EP elections from the official Facebook page of *Fidesz* were collected ($N=165$). The data collection spanned a five-week-period, starting from the month before the EP elections (23 April 2019) and lasting until a week after their end (2 June 2019). The first half-a-thousand comments that attracted the most likes and replies ($N=495$) given in response to the social media postings of the party were compiled. The genre of comments, typically brief and anonymous spontaneous reactions, was chosen for the present research as it allows for gaining insights into the genuine opinion, sentiments and feelings of the commenters. The compiled corpus included comments that were the most actively circulated in the discourse community in order to ensure that the data covered the most popular reactions, thus the most typical views among the supporters. The sample of comments was analysed using van Leeuwen's (2008) framework of discursive legitimization strategies. Van Leeuwen's comprehensive framework embraces four major categories (authorisation, moral evaluation, rationalisation and mythopoesis) with various further subdivisions. As several other subcategories emerged in the data, van Leeuwen's categorisation was extended. The supporters' comments were manually annotated according to the extended taxonomy.

The discursive constructions of legitimization that the supporters used in their social media activity showed a rich variety. The topoi of the supporters' discursive legitimations revealed that they promoted a system of values that involved norms both at the individual level and at the social level of their community. The identity of being Hungarian played a major role in the supporters' party choice. The discourse of the commenters shed light on what being Hungarian meant for the supporters, namely, belonging to the generic community of Hungarians, practicing Hungarian culture and upholding Hungarian traditions.

References

Mudde, Cas and Cristóbal R. Kaltwasser. (2017). *Populism: A very short introduction*. Oxford: Oxford University Press.

van Leeuwen, Theo. (2008). *Discourse and practice: New tools for Critical Analysis*. Oxford: Oxford University Press.

Data-driven learning in academic email communication

A comparative study in L1 and L2 Italian

Nicola Brocca (Universität Innsbruck, Austria)

Elena Nuzzo (Università Roma Tre, Italy)

Keywords: data-driven learning, formal email, pragmatic awareness, Italian L1/L2

Extensive pragmatic research has been conducted on students writing emails to academic staff in their first language (L1) (Biesenbach-Lucas 2007, 2009) and in intercultural communication contexts (Economidou-Kogetsidis 2011; Saviç 2018). This research extends to Italian L1 and L2 contexts, as explored by Andorno (2014), Clemenzi (2020) and Pagliara (2019, 2020). These studies have consistently observed that students often unintentionally make poor decisions related to communicative purpose or language usage, leading to unfavorable assessments of the sender's personality. Notably, research by Luijkx et al. (2022) highlights that receivers judge pragmatic infelicities in emails from non-native speakers as stringently as those from native speakers.

This contribution addresses these challenges by exploring the impact of a Data-Driven Learning (DDL) approach within a telecollaborative setting. The aim is to enhance understanding of pragmatic aspects in formal email writing in Italian for both native and non-native speakers. DDL promotes students' analysis of language using corpus data, encouraging a process of inquiry, investigation and discovery with authentic materials, thereby fostering greater autonomy in learning the target language (Boulton & Vyatkina 2021). Additionally, DDL is designed to develop lifelong, self-directed researchers – a crucial trait for professionals in a constantly evolving world.

The study introduces a 4-week module conducted in the Spring Semester of 2023, focusing on formal email writing. It involved eight monolingual, non-native pre-service teachers from the University of Innsbruck and their peers at the University of Roma Tre. The module incorporates task-based language teaching and corpus-based analyses using the corpus (Brocca 2022), focusing on the observation of pragmatic elements such as formulaic greeting forms in emails from both native and nonnative speakers.

Employing a mixed-methods approach, this study used a pre-post design questionnaire to assess pragmatic judgments of email components using a Likert scale, complemented by open follow-up questions to evaluate metapragmatic competencies. Results indicated a statistically significant improvement in non-native pre-service teachers, with minimal enhancement observed in native students. The study also revealed that positive corpus evidence bolstered awareness of appropriate form usage; for example, the closing formula "cordialmente" (kind regards) was recognized as appropriate with higher frequency post-intervention. Conversely, negative evidence had a lesser impact on identifying inappropriate usage; for instance, post-intervention, the expression 'Sig.re' (Mr.) followed by the professor's name was not more frequently recognized as inappropriate. Crucially, the findings suggest an increase in students' autonomy in investigating and

analyzing pragmatic issues, underlining the potential of DDL in enhancing pragmatic awareness for both native and non-native learners.

References

- Andorno, Cecilia. (2014). Una semplice informalità? Le e-mail di studenti a docenti universitari come apprendistato di registri formali. In *Lingue in contesto. Studi di linguistica e glottodidattica sulla variazione diafasica*. Massimo Cerruti M., Elisa Corino and Cristina Onesti (eds.) Dell' Orso, 1-20.
- Biesenbach-Lucas, Sigrun. (2007). Students writing emails to faculty: An examination of e-politeness among native and non-native speakers of English. *Language Learning and Technology* 11(2): 59-81.
- Biesenbach-Lucas, Sigrun. (2009). Little words that could impact one's impression on others: Greetings and closings in institutional emails. In *Little Words: Their History, Phonology, Syntax, Semantics, Pragmatic and Acquisition*. Ronald P., Héctor Campos and Donna Lardiere (eds.) Georgetown University, 183-195.
- Boulton, Alex and Nina Vyatkina. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology* 25(3): 66-89. <http://hdl.handle.net/10125/73450>.
- Clemenzi, Luca. (2020). Mettiamoci a scrivere. *L'esperienza del laboratorio LISA!* Dota M., Polimeri G. Prada M., Scrivere (eds.) oggi all'Università. Italiano LinguaDue 12:(1). <http://hdl.handle.net/2067/43609>
- Economidou-Kogetsidis, Maria. (2011). "Please answer me as soon as possible": Pragmatic failure in nonnative speakers' e-mail requests to faculty. *Journal of Pragmatics* 43(13). 10.1016/j.pragma.2011.06.006
- Economidou-Kogetsidis, Maria. (2023). "The language is very formal and appropriate": L2 learners' in/appropriateness evaluations and metapragmatic judgments in student-faculty emails. *Journal of Pragmatics* 217: 17-32.10.1016/j.pragma.2023.09.001
- Luijckx Antoinette, Gerritsen Marinel and Margot van Mulken. (2022). The importance of raising teachers' and students' awareness of pragmatics in German second language writing: A study of the effect of grammatical and lexical errors compared to pragma-linguistic infelicities. *Language Awareness* 31(1): 137-154. 10.1080/09658416.2021.2012479
- Pagliara, Francesca. (2019). La codifica pragmalinguistica dell'atto di richiesta nelle e-mail degli studenti universitari italiani. *Studi AltLA 9: Lingua in contesto. La prospettiva pragmatica*. Elena Nuzzo and Ineke Vedder (eds.): 149-168. http://www.aitla.it/images/pdf/StudiAltLA9/009_Pagliara.pdf
- Pagliara, Francesca. (2020). L'atto della richiesta nelle e-mail inviate dagli studenti universitari ai docenti: un confronto tra parlanti nativi e non nativi. *La scrittura per l'apprendimento dell'Italiano L2*. Roberta Grassi (eds.) Franco Cesati, 225-254.
- Savić, Milica. (2018). Lecturer perceptions of im/politeness and in/appropriateness in student e-mail requests: A Norwegian perspective. *Journal of Pragmatics* 124: 52-72. 10.1016/j.pragma.2017.12.005

Politicians VS LLMs

A corpus analysis on implicit persuasive techniques

Nicola Brocca (Universität Innsbruck, Austria)

Davide Garassino (Zürcher Hochschule für Angewandte Wissenschaften, Switzerland)

Viviana Masia (Università Roma Tre, Italy)

Keywords: political language, LLMs, implicit communication, manipulation

Human communication aims to influence beliefs “by design” (Mercier & Sperber 2017), but nonbona fide true (i.e. questionable or false) content can turn persuasion into manipulation (Reboul 2017). In particular, it is well recognized that certain linguistic strategies, such as implicit communication, may convey manipulative content effectively (Sbisà 2007). Lately, the worldwide success of Large Language Models (LLMs) has raised concern, as they can be easily exploited for manipulation (Burtell & Woodside 2023; Goldstein et al. 2023a, b). This is even more concerning since distinguishing between human and model-generated news is challenging for users and machines alike (Kreps et al. 2022; Chaka 2023). This study aims to explore the connection between implicit meaning and potentially manipulative content, with a focus on political communication, known for containing non-bona fide true implicit content (Lombardi Vallauri & Masia 2014; Garassino et al. 2019, 2022). Specifically, we delve into non-bona fide true presuppositions – conveyed by means of different triggers, such as definite descriptions (ex. *The black cat*), factive predicates (ex. *It’s strange she could not be here*), change-of-state verbs (i.e. John *stopped* telling lies), among others, as well as vague expressions whose meaning is likely to be underdetermined in certain contexts of language use (as is the case of some qualitative adjectives or common nouns).

In this investigation, we compare two small-sized corpora. The first is constructed from authentic speeches delivered by Italian, French and Spanish politicians, while the second consists of ChatGPT generated speeches that simulate those of the same politicians. Each corpus contains approximately 50,000 tokens and has been manually annotated.

To create the first corpus, we selected six politicians known for their diverse political views: Emmanuel Macron and Marine Le Pen from France, Giorgia Meloni and Nicola Zingaretti from Italy and Pedro Sánchez and Santiago Abascal Conde from Spain. We chose two speeches per politician, delivered during recent political rallies. The corresponding chatbot versions, created using ChatGPT4, constitute the second corpus.

This exploratory study poses several research questions:

- (a) How similar are the chatbot versions of politicians to real French, Spanish and Italian politicians?
- (b) Notably, is the frequency of presuppositions and vagueness strategies comparable in original texts and in ChatGPT-generated texts?

- (c) Is the discourse function (i.e., attack, praise, stance-taking, etc., cf. Garassino et al. 2019, a.o.) of presuppositions and vagueness strategies in Chat GPT's texts similar to that performed by the same strategies in real politicians' texts?

As of now, comparative analyses of real and automatically generated texts are part of a newly emerging research field that views Large Language Models (LLMs) as highly influential tools affecting how we use and comprehend language. Therefore, a corpus-based analysis should be meticulously conducted to gain a deeper understanding of this emerging phenomenon and to avoid succumbing to manipulative interactions.

References

- Burtell, Mathew and Thomas Woodside. (2023). *Artificial influence: An analysis of AI-driven persuasion*. 10.48550/arXiv.2303.08721
- Chaka, Chaka. (2023). Detecting AI content in responses generated by ChatGPT, YouChat and Chatsonic: The case of five AI content detection tools. *Journal of Applied Learning and Teaching* 6(2). 10.37074/jalt.2023.6.2.12
- Garassino, David. (2023). Measuring implicit communication. Some (tentative) proposals for a corpusbased pragmatics. Talk at the IMPAQTS Conference, Rome, 27-28.04.2023.
- Garassino, David, Brocca, Nicola and Viviana Masia. (2022). Is implicit communication quantifiable? A corpus-based analysis of British and Italian political tweets. *Journal of Pragmatics* 194: 9-22.
- Garassino, David, Masia, Viviana and Nicola Brocca. (2019). Tweet as you speak. The role of implicit strategies and pragmatic functions in political communication: Data from a diamesic comparison. *Rassegna Italiana di Linguistica Applicata* 2(3): 187-208.
- Goldstein, Josh A., Chao, Jason, Grossman, Shelby, Stamos, Alex and Michael Tomz. (2023a). *Can AI write persuasive propaganda?* 10.31235/osf.io/fp87b
- Goldstein, Josh A., Sastry, Girish, Musser, Micah, DiResta, Renee, Gentzel Matthew and Katerina Sedova. (2023b). Generative language models and automated influence operations: Emerging threats and potential mitigations. Stanford Internet Observatory, OpenAI and Georgetown University's Center for Security and Emerging Technology. <https://fsi.stanford.edu/publication/generative-language-models-andautomated-influence-operations-emerging-threats-and> (06.05.2023)
- Kreps, Sarah R., McCain, Miles and Miles Brundage. (2022). All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. *Journal of Experimental Political Science* 9(1): 104- 117.
- Levinson, Stephen. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Lombardi Vallauri, Edoardo. (2019). *La lingua disonesta*. Bologna: Il Mulino.
- Lombardi Vallauri, Edoardo and Viviana Masia. (2014). Implicitness impact: Measuring texts. *Journal of Pragmatics* 61: 161-184.
- Mercier, Hugo and Dan Sperber. (2017). *The enigma of reason*. Cambridge: Harvard University Press.
- OpenAI. (2021). ChatGPT (Version GPT-3.5) [Computer software]. <https://openai.com/>(06.05.2023)
- Pinker, Steven, Nowak, Martin A. and James J. Lee. (2008). The logic of indirect speech. *Proceedings of the National Academy of Sciences USA* 105: 833-838.
- Reboul, Anne. (2017). *Communication and cognition in the evolution of language*. Oxford: Oxford University Press.
- Sbisà, Marina. (2007). *Detto non detto. Le forme della comunicazione implicita*. Roma-Bari: Editori Laterza.

Language ideologies and identities of new speakers of Scots

A corpus-assisted discourse study

Linda Bruce (The Open University, United Kingdom)

Keywords: language ideologies, identity construction, Scots, new speakers, CADS

Scots is Scotland's largest minoritised language, with around 1.5 million speakers (National Records of Scotland 2013). From the 17th century onwards Scots underwent a process of dialectalisation and today's speakers tend to select forms in their everyday discourse that index a broad range of social meanings (Eckert 2008), within a "multi-dimensional sociolinguistic variation space" (Maguire 2012: 55). Although stigmatising language ideologies continue to circulate, painting Scots as a non-standard dialect of English, 'slang', or associated with rural, uneducated communities, more recently, a language revitalisation movement has developed, viewing Scots as a linguistic and cultural asset for modern-day Scotland. Since 2001, Scots has been protected and promoted through the European Charter for Regional or Minority Languages (Council of Europe 1992); it is increasingly present in Scottish schools and is more visible on social media and in broader cultural contexts.

However, no scholarly attention has so far been paid to new speakers of Scots, i.e. "language users who have typically made the conscious choice to acquire the minoritized language ... through individual efforts rather than ... through intergenerational transmission" (Hornsby 2022: 1). Using a corpus approach to discourse, this study investigates language ideologies of a group of new speakers of Scots in Scotland, as well as the ways they construct their identities in relation to their use of Scots and other languages. This research seeks to fill a gap in the scholarship concerning new speakers of minority languages and to help inform Scots language policymakers about the motivations, opinions and learning needs of this under-researched group of Scots users.

As a new speaker of Scots myself, I conducted semi-structured topic-based one-to-one interviews with 25 new speakers of Scots, as well as a discussion group with four traditional speakers working in Scots language teaching and support. I transcribed and POS-tagged the resulting data, creating a 220,000-word corpus of the speech of new speakers of Scots (CONSOS) and a supplementary corpus of Scots language professionals (SCOTSPROFS). Following Jaworska and Kinloch's framework for CADS with multiple data sets (2018), I identified candidate key items (Gabrielatos 2018), then classified them into semantic domains including *actors, emotions, ethnicity and nation, evaluation, language, ownership, self and identity*. Collocation patterns for key items were explored and discourse analysis carried out on examples relevant to the research question "How similar/different are new speakers of Scots, in terms of (1) language ideologies and identity constructions and (2) perceived differences between new and traditional speakers of Scots?"

This study highlights three prevalent language ideologies, which I have named as (1) *Language labels matter*; (2) *Language hierarchy in Scotland*; and (3) *The hierarchization of Scots speakers*. These ideologies are constructed, repeated, challenged and/or rejected by the

interviewees. The participants also intersubjectively construct their identities through discourse (Bucholtz & Hall 2010), taking up positions such as Scots language experts/enthusiasts, advocates and activists, insiders/outsideers, fakes, rebels, enablers, mediators and creative experimenters.

References

- Bucholtz, Mary and Kira Hall. (2010). Locating identity in language. In *Language and Identities*. C. Llamas & D. Watt (eds.) Edinburgh: Edinburgh University Press, 18-28.
- European Charter for Regional or Minority Languages, (1992). <https://www.coe.int/en/web/europeancharter-regional-or-minority-languages/text-of-the-charter>
- Eckert, Penelope. (2008). Variation and the indexical field 1. *Journal of Sociolinguistics* 12(4): 453-476.
- Gabrielatos, Costas. (2018). *Keyness analysis: Nature, metrics and techniques, Corpus Approaches to Discourse: A Critical Review*. Charlotte Taylor and Anna Marchi (eds.) London: Routledge, 225-258.
- Hornsby, Michael. (2022). 'Authentic' Language as a Contested Concept in Brittany. *Adeptus* (19).
- Jaworska, Sylvia and Karen Kinloch. (2018). Using multiple data sets. *Corpus approaches to discourse: A critical review*. Charlotte Taylor and Anna Marchi (eds.) London: Routledge, 110-129.
- Maguire, Warren. (2012). English and Scots in Scotland. *Areal features of the Anglophone world*. De Gruyter Mouton, 53-78.
- National Records of Scotland. (2013). Scotland's Census. <https://www.scotlandscensus.gov.uk/webapi/jsf/tableView/tableView.xhtml> (Retrieved 7 November 2023)

Mental health and illness in the Italian press

A corpus-assisted study

Tessa Cancio Pastor (Università di Pavia, Italy)

Chiara Zanchi (Università di Pavia, Italy)

Keywords: mental health, corpus-assisted discourse analysis, press discourse, Italian

The World Health Organization defines health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity”. Despite this definition, mental health discourse undergoes stigmatization and several studies have highlighted media’s crucial role in its propagation (Klin & Lemish 2008; Aragonès et al. 2014; Ma 2017; Atasanova et al. 2018; Brookes et al. 2022; Price 2022). Nevertheless, contributions on Italian are scarce and limited to specific conditions, such as schizophrenia (Carpiniello et al. 2007; Magliano et al. 2010; Pingani et al. 2018). This work contributes to covering this gap, providing a corpus-assisted analysis of how mental health and its conditions are represented in the Italian press.

In line with Price (2022), we aim to answer the following questions: Has the health-illness continuum shifted towards the negative pole? How is reference to illness expressed and linked to stigmatization? To answer these questions, we collected a corpus of 350 articles (236,872 tokens) published between 2007 and 2023 in three major Italian newspapers (La Repubblica, Il Corriere, La Stampa), using the software BootCaT (Baroni & Bernardini 2014). The keywords used to create the corpus cover general mental health discourse and specific conditions (depression, schizophrenia, eating disorders, bipolar disorder, personality disorders). The analysis involved calculating frequencies, collocates and manual annotation of concordances using AntConc (Anthony 2023) and SketchEngine (Kilgarriff et al. 2014).

The occurrences of *salute mentale* (mental health) were tagged as positive, neutral or negative depending on their context (5 words to the left and to the right). For example, if in the context there was mention of illness, hospitalization or crime, the occurrence was tagged as negative, whereas if there was mention of promotion or support as positive. Secondly, for each mental condition, we tagged occurrences of referential expressions as “identity first” if they contain nominalizations (i.e., “the *depressed*”), or as “person first” if mental conditions are expressed through modifiers (i.e., “person *suffering from depression*”).

The results show that *salute mentale* more frequently occurs than *malattia mentale* (mental illness) and *benessere* (wellbeing); most occurrences of *salute mentale* showed negative orientation (50%) followed by neutral (27%) and positive (23%). Among the top collocates, there were both neutral (*dipartimento*, “department”) and negative (*problemi* “problems”) words. Through SketchEngine’s Thesaurus function great overlap was also shown between *salute* and *malattia*. Compared to Price’s (2022) results, the situation in the Italian media seems to be more nuanced: the semantic shift seems milder and *salute* and *malattia* frequently share their contexts of occurrence. Concerning specific conditions, we found that “identity first” reference was mostly used for depression, followed by eating

disorders and schizophrenia. This type of reference, if used for personality disorders, schizophrenia and, to a limited extent, depression, was associated to violence and crime. However, the use of “identity first” expressions is not predominant: this could be due to the publication of guidelines for the representation of mental illness/health by journalistic associations (so-called Carta di Trieste 2011) as well as the increased sensitivity and exposure following the Covid-19 pandemic.

References

- Anthony, Laurence. (2023). AntConc (Version 4.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>
- Aragonès, Enric, López-Muntaner, Judit, Ceruelo, Santiago and Joseph Basora. (2014). Reinforcing stigmatization: Coverage of mental illness in Spanish newspapers. *Journal of Health Communication* 19(11): 1248-1258, 10.1080/10810730.2013.872726
- Atanasova, Dimitrinka., Koteyko, Nelya, Brown, Brian and Paul Crawford. (2019). Mental health and the media: From illness to wellbeing. *Sociology Compass* 13(5). 10.1111/soc4.12678
- Baroni, Marco and Silvia Bernardini. (2004). BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC*.
- Bowen, Matt, Kinderman, Peter and Anne Cooke. (2019). Stigma: A linguistic analysis of the UK red-top tabloids press' representation of schizophrenia. *Perspectives in Public Health* 139(3): 147-152. 10.1177/1757913919835858
- Brookes, Gavin, Atkins, Sarah and Kevin Harvey. (2022). Corpus linguistics and health communication: Using corpora to examine the representation of health and illness. *The Routledge Handbook of Corpus Linguistics*, 615-628. 10.4324/9780367076399-43
- Carpiniello, Bernardo, Girau, Roberta and Maria Germana Orrù. (2007). Mass-media, violence and mental illness. Evidence from some Italian newspapers. *Epidemiologia e Psichiatria Sociale* 16(3): 251- 255. [10.1017/s1121189x00002359n](https://doi.org/10.1017/s1121189x00002359n)
- Carta di Trieste (2011) available at: https://www.news-forumsalutementale.it/public/Carta_di_Trieste_2011.pdf
- Kilgarriff, Adam, Baisa Vit, Bušta, Jan, Jakubiček, Milos, Kovář, Vojtech, Michelfeit, Jan, Rychlý, Pavel and Vit Suchomel. (2014). The Sketch Engine: ten years on. *Lexicography* 1, 7-36.
- Klin, Anat and Dafna Lemish. (2008). Mental disorders stigma in the media: Review of studies on production, content and influences. *Journal of Health Communication* 13(5): 434-449. 10.1080/10810730802198813
- Ma, Zexin. (2017). How the media cover mental illnesses: A review. *Health Education* 117(1): 90-109. 10.1108/he-01-2016-0004
- Magliano, Lorenza, Read, John and Riccardo Marassi. (2010). Metaphoric and non-metaphoric use of the term “schizophrenia” in Italian newspapers. *Social Psychiatry and Psychiatric Epidemiology* 46(10): 1019-1025. 10.1007/s00127-010-0274-3
- Pingani, Luca, Sampogna, Gaia, Borghi, Giulia, Nasi, Annamaria, Coriani, Sandra, Luciano, Mario, Galeazzi, Gian Maria, Evans Lacko, Sara and Andrea Fiorillo. (2018). How the use of the term “schizo*” has changed in an Italian newspaper from 2001 to 2015: Findings from a descriptive analysis. *Psychiatry Research* 270: 792-800. 10.1016/j.psychres.2018.10.076
- Price, Hazel. (2022). *The language of mental illness: Corpus linguistics and the construction of mental illness in the press*. Cambridge University Press.

“I am a heterosexual woman who likes to party! vs. she/they, I live life to the fullest and sometimes draw stuff”

A corpus analysis study of WSM and WSW’s self-presentation strategies on Tinder profiles

Clara Cantos Delgado (Universidad Complutense de Madrid, Spain)

Carmen Maíz-Arévalo (Universidad Complutense de Madrid, Spain)

Keywords: corpus analysis, online dating, gender, WSW, self-presentation

The focus of this study is set on the different self-presentation strategies used by women-seeking-men (WSM) and women-seeking-women (WSW) on Tinder profiles in the UK, which builds on prior corpusanalytical research on dating profiles (Baker 2018; Collins 2019). Data was collected through Tinderbotz, a scraping tool which gathered the information publicly displayed on Tinder profiles. Then, to ensure an ethical research praxis, it assigned each profile random identification numbers that guaranteed the anonymity of the users and deleted all possibly identifying information (Instagram account name, Snapchat account name). The resulting sample was constituted by two sub-corpora (WSMC and WSWC), containing a total of 405 profiles (average number of words=40) from women (18-24 y.o.) based in the UK. Sketch Engine was used to search for significant n-grams and keywords in both sub-corpora and chi-square tests were conducted to test for significance. Tinder’s affordances determine that gender identity is established by selecting it among the options available (man/ woman/ Beyond binary) and/ or by alluding to it in the “about me” section of the profiles. In this regard, more instances of “woman” were found in the WSMC while the WSWC contained self-made descriptions of gender identity in their bios (as seen, for instance by the specification of gender pronouns, e.g., she/her; they/them), which can be interpreted as greater gender identity awareness. The lemmas “university”, “work” or “place” were found with similar frequency in both corpora. However, in terms of language variety, the WSMC seemed more homogeneous than the WSWC, as the latter included more diversity in terms of nouns and adjective selection. Nevertheless, some tendencies were still salient. The WSMC included more mentions of social networks, identified as nonsmokers and mentioned interests such as drinking, fashion and nightlife more frequently than the WSWC. Contrastingly, lemmas like “art”, “life” and “foodie” had a higher frequency in the WSWC. In terms of adjectives, “social” was among the most frequent attributes mentioned in both corpora, but it was much more frequent in the WSMC than in the WSWC. Contrastingly, “open” appeared more frequently in the WSWC than in the WSMC. Finally, sexual references (“sexy”, “hot”) were found in the WSWC but not in the WSMC. Our analysis of n-grams showed a similar interest in engaging in relationships by WSW and WSM, but more frequent specifications of turndowns (e.g., “I don’t”) were found in the WSMC. Further studies are needed to determine the possible motivations for these differences (platform type, age, socioeconomic background, etc.) and to determine the consistency of these findings.

References

Baker, Paul. (2018). Sexuality. In *Studies in Corpus-based Sociolinguistics*. Eric Friginal (ed.) London: Routledge: 157–175.

Collins, Luke. (2019). Dating apps: A Tinder corpus. In *Corpus linguistics for online communication: A guide for research*. London: Routledge.

Online news readers' metaphorical conceptions of climate change

Ben Clarke (Universitas Regia Gothoburgensis, Sweden)

Keywords: climate change, Conceptual Metaphor Theory, newsreader comments, digital discourse

Comprehending climate change is difficult given its severity, complexity and far-reaching nature. Likely for this reason, studies applying Conceptual Metaphor Theory have been popular within discourse-analytic work on climate change communication. Such work to date has identified four prominent metaphorical conceptions for communicating climate change: as CONFLICT, as RELIGION, as ILLNESS and as JOURNEYING (Atanasova & Koteyko 2017; Currie & Clarke 2022; Woods et al. 2012). However, such work has predominantly focused on archetypical bricks-and-mortar institutions, such as parliamentary politics, scientific institutions, traditional news media and corporate industry. Among others, Boykoff (2011) and Stibbe (2012) have argued that there is a risk that hegemonic, neo-liberal discourses both: (i) dominate the climate change debate, naturalising such conceptualisations and their consequences and (ii) limit our understanding of climate change phenomena, precluding other ways that climate change is – and can be – discussed and comprehended.

Here, a 75-million-word corpus of newsreader comments on all the Guardian Online climate change articles (2009-2023) are analysed using a corpus-assisted discourse studies methodology. Such data are a unique, highly opinionated and informationally-rich digital genre (Ehret & Taboada 2021), free from strict institutional logics. This study reports findings for three analytical inquiries concerning metaphorical conceptualisations of climate change in the data: first, it assesses the presence and pragmatic characteristics of the aforementioned conceptual metaphors, finding JOURNEYING and CONFLICT to be most salient. Second, the presence of other conceptual metaphors not previously reported in the literature from prior studies (e.g. as MONEY) and their characteristics are noted. Third, given the polylogic (Marcoccia 2004) nature of this digital media, of the metaphorical conceptualisations identified, it is determined which are most challenged by fellow commenters (typically, CONFLICT metaphors) and how so discursively.

References

- Atanasova, Dimitrinka and Nelya Koteyko. (2017). Metaphors in Guardian online and mail online opinion-page content on climate change: War, religion and politics. *Environmental Communication* 11(4): 452–469.
- Boykoff, Maxwell T. (2011). *Who speaks for the climate? Making sense of media reporting on climate change*. Cambridge: Cambridge University Press.
- Currie, John S. G. and Ben Clarke. (2022). Fighting talk: The use of the conceptual metaphor climate change is conflict in the UK Houses of Parliament, 2015-2019. *Journal of Language and Politics* 21(4): 589-612.
- Ehret, Katharina and Maite Taboada. (2021). Characterising online news comments: A multi-dimensional cruise through online registers. *Frontiers in Artificial Intelligence* 4.
- Marcoccia, Michel. (2004). On-line polylogues: Conversation structure and participation framework in internet newsgroups. *Journal of Pragmatics* 36(1): 115-145.
- Stibbe, Arran. (2012). Ecolinguistics and globalization. In *The Handbook of Language and Globalization*. Nikolas Coupland (ed.) Malden: Wiley-Blackwell, 406-425.
- Woods, Ruth, Fernández, Ana and Sharon Coen. (2012). The use of religious metaphors by UK newspapers to describe and denigrate climate change. *Public Understanding of Science* 21(3): 323-339.

The discourses of vaccination across pseudoscience and conspiracy websites

Isobelle Clarke (Lancaster University, United Kingdom)

Yuze Sha (Lancaster University, United Kingdom)

Keywords: anti-science, keyword co-occurrence analysis, multiple correspondence analysis, multi-dimensional analysis, ATLAS.ti

Science improves our everyday lives, especially science that is aimed at safeguarding and protecting public health and safety, such as by improving the air we breathe and the water we drink (Carter et al. 2019). Because of science, food is safe and plentiful and diseases can be treated, cured, isolated and prevented from spreading (Siegel 2017). Science also anticipates threats to the environment and natural disasters, like hurricanes and storms (Carter et al. 2019). Although scientific advancements can be misused accidentally or for ill, science nevertheless has led to the development of new technologies, which have enhanced many individuals' quality of life to a level that could never have been expected previously (Siegel 2017). Yet, these advances in humanity's safety and prosperity are being significantly undermined and threatened by anti-science discourse and actions. In this talk, we focus on the discourses of a particular strand of anti-science; namely, anti-vaccination. Specifically, we use Keyword Co-occurrence Analysis (KCA) (Clarke et al. 2021) to uncover dimensions representing the most dominant patterns of co-occurring keywords across a corpus of 37,676 texts from pseudoscience and conspiracy websites that mention *vaccin**. We then use Atlas.ti to assist in the interpretation of our dimensions from multiple angles, including discourse, register, style and vaccine attitude. Our results demonstrate some of the persuasive strategies and repertoires used by those expressing anti-vaccination sentiment.

References

Carter, Jacob, Berman, Emily, Desikan, Anita, Johnson, Charise and Gretchen Goldman. (2019) The state of science in the Trump era: Damage done, lessons learned and a path to progress. *Center for Science and Democracy at the UCS*.

Siegel, Ethan. (2017). Humanity needs science to survive and thrive. <https://www.forbes.com/sites/startswithabang/2017/07/25/humanity-needs-science-to-survive-and-thrive/#279e490fe5a2> (Accessed 21/01/2020).

Clarke, Isobelle, McEney, Tony and Gavin Brookes. (2021). Multiple correspondence analysis, newspaper discourse and subregister: A case study of discourses of Islam in the British press. *Register Studies* 3(1): 144-171.

Environmental sustainability in the ESG reports of Chinese listed companies

A preliminary corpus-assisted analysis from an ecolinguistic perspective

Sergio Conti (Università Roma Tre, Italy)

Laura Locatelli (Università Ca' Foscari, Italy)

Daniele Brombal (Università Ca' Foscari, Italy)

You Pui Szeto (Università Ca' Foscari, Italy)

Keywords: corpus-assisted ecolinguistics, Chinese ESG, keyword analysis, environmental discourse

Amidst global concern for the escalating environmental crisis, Environmental, Social and Governance (ESG) reporting has emerged as a standard practice among listed companies, to record and communicate their commitment and actions towards sustainability. China makes no exception: indeed, the diffusion of ESG reporting among Chinese companies has followed a pattern similar to their Western counterparts. Besides providing investors and stakeholders with information, ESGs are a powerful discursive tool to shape companies' image, vis-à-vis mounting public pressure for corporate accountability (Yu & Bondi 2019).

In recent years, linguistic research has investigated the discursive dimension of Chinese ESGs, employing methods such as critical discourse analysis (Gong 2019), critical metaphor analysis (Yu 2020) and cross-cultural comparison (Huang 2019; Yu & Bondi 2017). Despite progress made, little effort has been made to incorporate an ecolinguistic approach (Stibbe 2021) in the analysis of environmental discourse in Chinese ESGs. We aim to fill this gap by means of a systematic, ecolinguistics-informed study of vocabulary choices and themes employed in Chinese ESG reports.

To this end, we collected a sample of ESG reports from companies listed in the "2021 Chinese Largest Companies per Revenue" (List A), mostly State-owned and those included in the "2021 Hurun China Private Companies SDG Readiness 100" (List B). Based on the different ownership of the companies in the two lists and since companies in List B comply with Sustainable Development Goals (SDGs) criteria, our assumption is that the two lists may reflect diverse attitudes toward sustainability. After eliminating redundancies from the two lists, we extracted the environmental sustainability section from each report. The final corpus contains 136 semantically-tagged reports (68 in each List), with a size of approximately 500,000 words.

Keyword analysis comparing List A and List B was conducted on Sketch Engine and #LancsBox X for words, multi-word terms and semantic tags. The results obtained by triangulating keyword analysis and manual clustering of keywords with semantic tagging highlight different sustainability and political discourses across the two subcorpora. Reports in List A resonate with the Party-State agenda, setting a general blueprint towards the attainment of overarching national goals. The discourse of progress towards these goals appears constructed by making reference to overarching principles and archetypal cases,

rather than insisting on monitoring and accurate metrics. Conversely, reports in List B are more down to earth and tangible, in that they highlight the impacts of corporate operations and mention criteria to monitor them, alongside with measures to mitigate potential harm. By doing so, they also appear less prone to the risk of erasing contentious environmental issues. From an ecolinguistic perspective, these findings shed light on the building blocks of diverse environmental discourses in ESG reporting and the way corporate environmental discourse may be normalized. On a more political note, they highlight two different approaches that currently coexist in China's sustainability arena: on the one hand, an ideology of sustainability rooted in official, wide-ranging discourses (List A), which by constant reiteration may become interiorized by citizens and power holders alike, strengthening at the same time the centrality of the PartyState; on the other hand (List B), the provision of specific criteria that may enable the public (or part of it, e.g. the companies' shareholders) to hold companies to account for their progress (or lack thereof) in meeting sustainability targets.

References

- Gong, Heng. (2019). A corpus-based critical ecological discourse analysis of corporate annual environmental reports: China three gorges corporation as an example. MA diss. University of Helsinki.
- Huang, Minyan. (2019). Comparative studies of Chinese & English corporate ESG reports and revelations of C-E publicity translation. *Frontiers in Educational Research* 2(2): 106-111. 10.25236/FER.033015
- Stibbe, Arran. (2021). *Ecolinguistics: Language, ecology and the stories we live-by*. Abingdon & New York: Routledge.
- Yu, Danni. (2020). Corporate environmentalism: A critical metaphor analysis of Chinese, American and Italian corporate social responsibility reports. *IEEE Transactions on Professional Communication* 63(3): 244-258. 10.1109/TPC.2020.3012728
- Yu, Danni and Marina Bondi. (2017). The generic structure of CSR reports in Italian, Chinese and English: A corpus-based analysis. *IEEE Transactions on Professional Communication* 60(3): 273-291. <https://doi.org/10.1109/TPC.2017.2702040>
- Yu, Danni and Marina Bondi. (2019). A genre-based analysis of forward-looking statements in corporate social responsibility reports. *Written Communication* 36(3): 379-409. 10.1177/0741088319841612

Conceptualising climate change in the UK Houses of Parliament

A corpus-assisted discourse analytical approach

John Currie (Universitas Regia Gothoburgensis, Sweden)

Daniel Lees Fryer (Østfold University College, Norway & Universitas Regia Gothoburgensis, Sweden)

Ben Clarke (Universitas Regia Gothoburgensis, Sweden)

Keywords: climate change, UK parliamentary discourse, metaphors, appraisal patterns

Climate change is an issue which affects almost every area of human and nonhuman life. It is so woven into the fabric of society that fully comprehending or articulating the issue is challenging and potential solutions can appear irrational (Morton 2013; Lazarus 2009). Studying the language used to conceptualise climate change is crucial as a means of understanding how the issue itself, as well as ways of dealing with climate change, are perceived and acted on (e.g. Fløttum 2017: 7-8). Previous research examining climate change discourse while employing corpus approaches to discourse studies include analyses of news media (Gillings & Dayrell 2023), blogs (Fløttum, Gjesdal, Gjerstad, Koteyko, & Salway 2014), newsreader comments (Koteyko, Jaspal, & Nerlich 2013) and parliamentary debates (Currie & Clarke 2022).

Here, we build on existing research, focussing on how climate change is conceptualised in UK parliamentary discourse. Parliamentary discourse is, of course, crucial to study in the context of climate change, not only due to the social power that political decision-makers possess, but also as it is in parliament where the necessary legislative action can be taken to enact bold, society-level changes to deal with climate change (Currie & Clarke 2022; Willis 2017). The UK parliament has been relatively fast to act, compared with other countries, in terms of setting climate change targets, but clear policy to achieve these targets has been lacking (Willis 2019: 889).

We take a diachronic approach, exploring the extent to which discursive constructions of climate change have altered from when the term first began appearing regularly in the UK Houses of Parliament in the late 1980s and 1990s, towards the present day. This kind of research helps to better understand contemporary uses and understandings of climate change (Taylor 2021: 464). Analysis of verb collocates of the term 'climate change' in Hansard data (Hansard, n.d.) shows the use of active verbs including 'address', 'mitigate', 'deal with', as well as conflict metaphors, including 'tackle', 'combat' and 'fight' have been becoming more frequent over time as we move towards the present day, suggesting that UK political decisionmakers are increasingly discussing the need to take action against climate change. The increase of conflict metaphors can also be observed in the growing use of nouns such as 'fight' and 'battle' (e.g. 'the UK's battle against climate change'). Adjective collocates, analysed in terms of appraisal (Martin & White 2005), reveal that language used to describe climate change has been becoming more dramatic or serious in recent years, with terms such as 'catastrophic' and 'existential' increasing in use and becoming top-ranking collocates in 2019 and 2020, whereas terms such as 'important' – ranked highly in the early 2000s and

perhaps comparatively more neutral – drop off in use. Moreover, climate change is increasingly described as a ‘threat’ and a ‘challenge’ in the 2000s and 2010s. These findings could reveal that UK political decision-makers’ overall conceptualisation of climate change may well be moving in line with newly introduced terms such as ‘climate emergency’ and ‘climate crisis’.

References

- Currie, John and Ben Clarke. (2022). Fighting talk: The use of the conceptual metaphor CLIMATE CHANGE IS CONFLICT in the UK Houses of Parliament 2015-2019. *Journal of Language and Politics* 21(4): 589-612.
- Fløttum, K. (2017). Language and climate change. In *The role of language in the climate change debate*. Kjersti Fløttum (ed.) London: Routledge, 1-9.
- Fløttum, Kjersti, Gjesdal, A., Gjerstad, Oyvind, Koteyko, Nelya and Andrew Salway. (2014). Representations of the future in English language blogs on climate change. *Global Environmental Change* 29: 213-222.
- Gillings, Mathew and Carmen Dayrell. (2023). Climate change in the UK press: Examining discourse fluctuation over time. *Applied Linguistics*.
- Hansard. n.d. Hansard. <https://hansard.parliament.uk/> (Accessed 11 January 2024).
- Koteyko, Nelya, Jaspal, Rrusi and Brigitte Nerlich. (2013). Climate change and 'climategate' in online reader comments: A mixed methods study. *The Geographical Journal* 179(1): 74-86.
- Lazarus, Richard. (2009). Super wicked problems and climate change: Restraining the present to liberate the future. *Cornell Law Review* 94(5): 1153–1233.
- Martin, J. Sherwin and Peter White (2005). *The language of evaluation: Appraisal in English*. London: MacMillan.
- Morton, Timothy. (2013). *Hyperobjects: philosophy and ecology after the end of the world*. University of Minnesota Press.
- Taylor, Charlotte. (2021). Metaphors of migration over time. *Discourse & Society* 32(4): 463-481.
- Willis, Rebecca. (2017). Taming the climate? Corpus analysis of politicians’ speech on climate change. *Environmental Politics* 26(2): 212–231.
- Willis, Rebecca. (2019). The role of national politicians in global climate governance. *Environment and planning E: Nature and Space* 3(3): 885-903.

The discursive construction of the climate and health polycrisis in English, French and Spanish

Niall Curry (Manchester Metropolitan University, United Kingdom)

Gavin Brookes (Lancaster University, United Kingdom)

The climate crisis is arguably the greatest challenge of our time, impacting the future of our planet and all life inhabiting it. As a result, there has been a surge of interest in addressing this issue in corpus and discourse studies. Largely focusing on the English language, studies in the area have centred on a range of texts and contexts, revealing and critically analysing the different actors involved in the exacerbation of climate issues, e.g., climate sceptics, corporations, government, lobbyists. Likewise, different representations of climate issues have been addressed, spanning foci on extreme weather, deforestation and climate scepticism, for example. In many cases, these actors and representations are evoked to signal the impact of the climate crisis on the likes of the economy, human, animal and planetary health and food production. Among these areas of impact, the interconnectedness of climate and health crises is noteworthy (Bevitori & Russo 2023), with entities like the World Health Organisation declaring climate change not just an environmental issue but a health issue in and of itself. Yet, despite the evident links between climate and health issues, corpus-assisted discourse studies of climate and health crises have typically centred on the investigation of how such crises are independently discursively constructed. This paper addresses this dearth in both the study of the climate and health crises as a form of intersecting polycrisis and multilingual studies of this phenomenon.

This analysis draws on a corpus of parascientific communication from *The Conversation's* English, French and Spanish language websites with a view to understanding better how experts from different cultural backgrounds socially construct the climate and health polycrisis for non-specialist readers. The corpus is composed of every climate-themed blog published in *The Conversation* before January 2024, reflecting an evident *tertium comparationis* worthy of analysis (Curry 2021). Using keyword and concordance analysis, a number of framing activities are identified that demonstrate the myriad ways in which climate and health crises intersect and how this process can differ across cultures. These include: defining and contextualizing the polycrisis; representing cause and effect; and proposing solutions. When interrogated to determine variation in the cultural and epistemological construction of knowledge of the climate and health polycrisis, the findings highlight the crucial role of cultural and linguistic diversity in shaping responses to global crises. In concluding, we make a case for pluriversal approaches to knowledge production to address the complex challenges posed by global polycrises.

References

Bevitori, Cinzia and Katherine E. Russo. (2023). Environment, climate and health at the crossroads: A critical analysis of public policy and political communication discourse in the EU. In *Handbook of Political Discourse*. Piotr Cap (ed.) Edward Elgar Publishing, 328-344.

Curry, Niall (2021). *Academic writing and reader engagement: Contrasting questions in English, French and Spanish corpora*. London: Routledge.

Changing meanings in discourses around politics

Exploring the outrage culture keywords

Daria Dayter (Tampere University, Finland)

Keywords: political discourse, cancel culture, Trump, news discourse, persuasion

The value of using corpus-assisted discourse studies, an approach to the study of discourse based on the computer-aided analysis of large volumes of naturally occurring data, has been frequently demonstrated in research projects that attempt to make sense of societally relevant issues. The more corpus-oriented branch of this approach has specifically worked on the analysis of discourse keywords. Discourse keywords are lexical items that are frequent in the discourses under scrutiny, function as semantic nodes at which an in-depth analysis unfolds, cluster with other discourse keywords prominent in the discourse under scrutiny (Schröter et al. 2019; Taylor & Del Fante 2020). They also often concern controversially debated issues in public discourses: for example, “illegal immigrants” or “austerity” (cf. Schröter & Veniard 2016; Taylor 2017).

In this talk, I will approach the study of persuasive discourses on societally relevant issues by focusing on discourse keywords. In particular, I will report an analysis of the discourse keywords “cancel culture”, “woke” and “outrage” (culture), in a corpus of Anglophone news on the web (NOW corpus, english-corpora.org) over the period of 2015-2023. By tracing the relative frequencies of these discourse keywords in three-months increments I identify the periods of time when they are especially salient and demonstrate the parallels in their use, strongly associated with the landmark speeches by Donald Trump. I follow this with a manual discourse analysis of their collocates, similarly identified in three-months increments and show that the emerging semantic groupings demonstrate pejorative development over time. Finally, through manual analysis of random samples of concordances, I demonstrate that different wordforms of the same discourse keyword may have distinctly different usages: for example, while “cancel culture” has almost exclusively negative meaning post-January 2020, “to cancel someone” appears in positive context approximately half the time and refers to a morally just and fair action.

References

- Schröter, Melani and Marie Veniard. (2016). Contrastive analysis of key words in discourses: Intégration and integration in French and German discourses about migration. *International Journal of Language and Culture* 3(1): 1-33. 10.1075/ijolc.3.1.01sch
- Schröter, Melani, Veniard Marie and Charlotte Taylor. (2019). A comparative analysis of the keyword multicultural(ism) in French, British, German and Italian migration discourse. In *Migration and Media: Discourses about Identities in Crisis*. Andreas Musolff and Lorella Viola (eds.) Amsterdam: Benjamin, 13-44.
- Taylor, Charlotte. (2017). Togetherness or othering? In *Representing the Other in European media discourses*. Jan Chovanec and Katarzyna Molek-Kozakowska (eds.) Amsterdam: Benjamins, 55-79.
- Taylor, Charlotte and Dario Del Fante. (2020). Comparing across languages in corpus and discourse analysis: Some issues and approaches. *Meta* 65(1): 29-50.

Vaccine and vaccination in parliamentary discourse in Brazilian Portuguese during COVID-19

Analysis of processes and appraisal

Rodrigo Esteves de Lima-Lopes (Universidade Estadual de Campinas, Brazil)

Keywords: discourse analysis, Systemic-Functional Linguistics, parliamentary discourse, evaluativity, transitivity, Brazilian Portuguese

This presentation discusses the initial results of processing the 'BRPoliCorpus' (Lima-Lopes 2022), a corpus of Brazilian political language. The corpus has been under development since 2022 and includes a database containing all inaugural speeches of Brazilian presidents from 1889 to 2022. It has expanded within a collaboration between UNICAMP and TU-Darmstadt, focusing on parliamentary discourse and candidate's government projects. The data analysed in this presentation is restricted to parliamentary debates collected during the COVID-19 pandemic, as data related to government programs are still undergoing collection and encoding processes. I will reflect on the challenges of data collection and organisation, as well as conduct an exploratory analysis of relational processes related to plenary sessions discussing terms such as *vacina* ('vaccination') and *vacina* ('vaccine') during this period. Despite data availability to the public due to Brazilian government regulations, these data need to be compiled in a structured manner and have their sparse metadata organised. Additionally, the data is provided through dynamic links, complicating mass scraping. Regarding the chosen theme, the interest in discourse arises from the evident opposition between pro and anti-vaccine discourses established in the Brazilian Congress during the period. Data scraping was made possible by reconstructing links through Python scripts (Python 2023), which reconstructed the original URLs. Data analysis was conducted using the R programming language (R Core Team 2021), with tagging, concordance and placement packages. Thus, processes served as research nodes, observing nouns and adjectives related to them. The theoretical framework of this work is on Systemic-Functional Linguistics (Halliday 2004), specifically in the systems of transitivity and appraisal (Martin & White 2005; Thompson & Alba-Juez 2014). The transitivity system is a tool for understanding the quantum of transformation and representation of human experience, while the appraisal system is a tool for understanding our position towards someone or an idea, how we evaluate it and to what extent we affiliate with it. This research seeks to analyse how processes and their arguments are chosen when endorsing, defining, denying and assessing the concepts of vaccine and vaccination within the corpus. The approach is comparative, examining whether evaluative and representational patterns establish themselves differently depending on the ideological affiliations of politicians. The results show consistent lexical-grammatical patterns aligned with the discursive positioning of parliamentary representatives.

References

- Halliday, Michael Alexander Kirkwood. (2004). *An introduction to functional grammar*. London: Arnold. 10.4324/9780203431269
- Lima-Lopes, Rodrigo Esteves de. (2022). BrPoliCorpus: Brazilian political corpus. <https://github.com/rll307/BrPoliCorpus.git>
- Martin, J. R. and P. R. R. White. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave Macmillan.
- Python. (2023). <https://www.python.org>
- R Core Team. 2021. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. <https://www.R-project.org>
- Thompson, Geoff and Laura Alba-Juez. (2014). Evaluation in context. *Pragmatics & beyond New Series 242*. Amsterdam: John Benjamins Publishing Company.

Re-presentation

A corpus-assisted semantic profile of the noun representation in CADS studies on newspaper discourse

Dario Del Fante (Università degli studi di Ferrara, Italy)

Keywords: newspaper discourse, CADS, representation and language

The existing literature on newspaper discourse from a CAD (Critical Discourse Analysis) or CADS (Corpus-assisted Discourse Studies) perspective is extensive and there is a large volume of published studies that mention the noun representation in their titles or who put the representation in the press as the main object of investigation. As argued by Partington, “the adoption of this particular term, in preference to other possibilities presumably deemed less functionally appropriate (...), is not without some interesting repercussions” (2015:1).

Generally, these works focus on the use of language to represent a social phenomenon like migration (KhosraviNik 2009; KhosraviNik et al. 2012), Islam (Baker et al. 2013), Europe (Marchi & Taylor 2008), homelessness (Schneider et al. 2010; Zufferey 2014), obesity (Brookes and Baker 2021) and many other events. However, except for some works such as Mitchell (1991) for literary studies, Fowler (1991), Hall (1997) and Partington (2015) for news media studies, Van Dijk (2002) concerning political discourse and, more recently, the work by Brookes and Baker (2021: 2-7), the explanation of what the act of representation means is often not directly and overtly addressed. In general, there are references to concepts such as ‘discursive representation’, ‘representation of social actors’ or ‘representation’, but there is no explicit explanation of what representation precisely means and how it works in language. However, the fact that the notion of representation is not often clearly defined in many works (generally for a matter of space) does not mean that the notion is taken for granted or the implications are not considered. For this reason, it would be interesting to investigate how the notion of representation is presented. On the basis that the meaning of an expression derives from its use in context, which is intended as all its manifestations through which the expression itself acquires its meaning, the aim of this work is to give a corpus-assisted discourse perspective on the use of the lexical item representation in the academic discourse focused on the news media. A written English corpus of approximately 200.000 words has been compiled by collecting open-access papers (or papers whose I have access to) whose title contains the words representation and press OR representation and newspaper. The papers have been retrieved through Google Scholar.

The methodology has been composed of two steps:

1. A priori definition of the meaning of representation from the main literature on the topic;
2. Analysis of the semantics of representation by investigating the collected corpus in order to formulate the a posteriori definition of the meaning of representation.

The general aim of this work is to understand whether there is a common agreement among researchers on the role played by the notion of representation in the press given the centrality of such a concept. This study does not pretend to give a full account of the use of representation in general. Its scope is narrowed around such academic research interested in examining how language is used to present reality through newspapers.

References

- Baker, Paul, Gabrielatos, Costas and Tony McEnery. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge University Press.
- Brookes, Gavin and Paul Baker. (2021). *Obesity in the news: Language and representation in the press*. Cambridge University Press. 10.1017/9781108864732
- Fowler, Roger. (1991). *Language in the news*. Routledge.
- Hall, Stuart. (1997). *Representation: Cultural representations and signifying practices*. London: Sage.
- KhosraviNik, Majid. (2009). The representation of refugees, asylum seekers and immigrants in British newspapers during the Balkan conflict (1999) and the British general election (2005). *Discourse & Society* 20(4): 477-498. 10.1177/0957926509104024
- KhosraviNik, Majid, Krzyżanowski, Michal and Ruth Wodak. (2012). Dynamics of representation in discourse: Immigrants in the British press. In *Migrations: Interdisciplinary Perspectives*. Michi Messer, Renee Schroeder and Ruth Wodak (eds.) Springer Vienna, 283–295. 10.1007/978-3-7091-0950-2_26
- Marchi, Anna and Charlotte Taylor. (2009). Establishing the EU: The representation of Europe in the press in 1993 and 2005. In *Corpora: Pragmatics and discourse*. Brill, 203–226.
- Partington, Alan. (2015). Corpus-assisted comparative case studies of representations of the Arab world. In *Corpora and Discourse Studies*. Paul Baker and Tony McEnery (eds.) Palgrave Macmillan UK, 220-243. 10.1057/9781137431738_11
- Schneider, Barbara, Chamberlain, Kerry and Darrin Hodgetts. (2010). Representations of homelessness in four Canadian newspapers: Regulation, control and social order. *J. Soc. & Soc. Welfare* 37: 147.
- Van Dijk, Teun. (2002). Political discourse and political cognition. *Politics as Text and Talk: Analytic Approaches to Political Discourse* 203: 203–237.
- Zufferey, Carole. (2014). Questioning representations of homelessness in the Australian print media. *Australian Social Work* 67(4): 525–536. 10.1080/0312407X.2013.842604

Lexical Priming Theory

Evolution, evaluation, extension

Eugenia Diegoli (Università di Bologna, Italy)

Alan Partington (Università di Bologna, Italy)

Keywords: Lexical Priming, Michael Hoey, semantic association, evaluation, evaluative cohesion

Michael Hoey's (2005) Lexical Priming (LP) theory was revolutionary in a number of ways. It was the first lexis-driven theory of language to combine large quantities of authentic 'usage-data' (in corpora) and psycholinguistic/cognitive research to explain the functioning, acquisition and (re)production of language in appropriate contexts. At the same time, it married the study of micro-contexts, typical of CL, with the macro-textual vision of the experienced discourse analyst that Hoey was.

According to the 2005 version of LP theory, during any encounter with a word (or a combination of words), we subconsciously note the words and the structures it commonly occurs with (its collocates and colligations), the semantic sets with which it normally co-occurs (its semantic associations) and the communicative functions it conveys (its pragmatic associations). We also note discourse patterns that operate within a longer distance from the word in focus (such as problem-solution, cause-effect, etc.). The more frequent the encounters with the word the more fixed in memory these associations, that is, *primings*, become.

With the aid of a series of Hoey's last unpublished lecture slides (numbering 1,499, 2014-2017), which he requested his family leave to us to work on and the Japanese language written corpus of online communication Yahoo! Chiebukuro Data, we examine current LP theory, some of the ways it has evolved, some of its unresolved issues and add a suggestion or two of our own (Hoey himself stresses repeatedly that this is a theory in development).

Some issues that remain unresolved in the 2005 version include some confusion about the status of *word* as against the morpho-phoneme, as well as some scepticism about the Sinclairian notion of *lexical item* (Sinclair 2004). What is the precise distinction between *semantic* and *pragmatic* associations – why for instance is UNCERTAINTY pragmatic but UNEXPECTED semantic? If pragmatic meanings are codified and conventionalised in the collective minds of a community of speakers, do they not, as many argue, belong in the realm of semantics? And most significant of all is the lack of attention paid to evaluation (i.e., the appraisal of an entity as essentially good or bad; Hunston & Thompson 2000) at both the micro and macro (cohesion) levels of analysis.

Using Hoey's own favourite examples (*consequences, in broad daylight, NOT hear a word against* <someone or something>, <turn> *a blind eye to* <something bad>, etc.), we show how evaluative meanings are pervasive at all levels and primed in speakers' minds in the same way as the other phenomena Hoey pays attention to. We then add our own thoughts on evaluative cohesion (Partington 2017).

Finally, the most striking novelty in Hoey's late unpublished slides is his desire to have LP theory tested on languages other than English to accumulate evidence to corroborate the fact that LP is a universal theory of language (with local adaptations). This talk is part of a project to test LP on a particular discourse type of Japanese (of which the first speaker is a fluent speaker) in an attempt to carry out Michael Hoey's wishes.

References

- Hoey, Michael. (2005). *Lexical priming: A new theory of words and language*. Milton: Taylor & Francis.
- Hunston, Susan and Geoff Thompson. (2000). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press.
- Partington, Alan. (2017). Evaluative clash, evaluative cohesion and how we actually read evaluation in texts. *Journal of Pragmatics* 117: 190-203.
- Sinclair, John. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Multimodal meaning-making

Showings in video-mediated conversations

Stefan Diemer (Hochschule Trier, Germany)

Keywords: corpus-assisted discourse analysis, video-mediated conversations, grammar, embodiment, multimodality

This paper analyzes how spoken utterances and embodied behavior combine to create meaning in showings (the display or foregrounding of objects), in video-mediated conversations. The analysis is based on two datasets, ViMELF (2018) and TaCoCASE (2023). Both corpora contain dyadic conversations between previously unacquainted partners via Skype. Gestures were separately transcribed, analyzed and categorized to permit both quantitative and qualitative analysis. The corpora are available online via WebCorpLSE (2024).

The role of gestures in the creation of meaning has been investigated extensively; less so in videomediated conversations, where the affordances of the medium restrict the scope of gestures to some extent. Previous studies in face-to-face environments (Goodwin 2010) and video-mediated settings (Licoppe 2017) indicate that showings are used to supplement lexis, but that the visual channel can also create a common focus of attention, organize the showing sequence and initiate interactive assessment (Rosenbaun & Licoppe 2017; Mondada 2019).

Participants in the two datasets use showings to introduce objects into the conversation by holding them into the pickup field of the camera. Showings can supplement the lexical designation for the item in question, either simultaneously, or consecutively. They also can replace the verbal level partially or completely. If the object is not available, showings can make use of forced perspective shift by moving the camera or, alternatively, imitating the lexical referent through the use of gestures and facial expressions. Showings also provide an opportunity to semantically frame the lexical item, e.g. by introducing humor or irony, particularly in combination with other nonverbal behavior such as laughter or pausing, opening up the possibility of interactive assessment.

The study illustrates the fundamental nature of language as an embodied activity. Showings in videomediated communication are 'environmentally coupled gestures' (Goodwin 2007: 197) that combine multiple semiotic fields and constitute a complex collaborative activity.

References

- Goodwin, Charles. (2007). Environmentally coupled gestures. In *Gesture and the Dynamic Dimension of Language*. Susan D. Duncan, Justine Cassell and Elena T (eds.) Amsterdam: John Benjamins, 195-212.
- Goodwin, Charles. (2010). Things and their embodied environments. In *The cognitive life of things: Recasting the boundaries of the mind*. Colin Renfrew (eds.) Cambridge: McDonald Institute for Archaeological Research, 103-120.
- Licoppe, Christian. (2017). Showing objects in Skype video-mediated conversations: From showing gestures to showing sequences. *Journal of Pragmatics* 110: 63-82.
- Mondada, Lorenza. (2019). Practices for showing, looking and videorecording: The interactional establishment of a common focus of attention. In *Embodied activities in face-to-face and mediated settings: Social encounters in time and space*. Elisabeth Reber and Cornelia Gerhardt (eds.) Berlin: Springer, 63-104.
- Rosenbaun, Laura and Christian Licoppe. (2017) Showing 'digital'objects in web-based video chats as a collaborative achievement. *Pragmatics* 27(3): 419-446.
- TaCoCASE. (2023). Transatlantic component of the CASE project. Birkenfeld: Trier University of Applied Sciences. [<http://umwelt-campus.de/case/TaCoCASE>].
- VIMELF. (2018). Corpus of video-mediated English as a lingua franca conversations. Birkenfeld: Trier University of Applied Sciences. Version 1.0. The CASE project [<http://umwelt-campus.de/case>].
- WebCorpLSE. 2024. Linguist's search engine. Compilation: Research and development unit for English Studies, Birmingham City University. <https://www.webcorp.org.uk/wcx/lse/>

Evaluative cohesion and the values of delegitimization

Alison Duguid (Università di Siena, Italy)

Dario Del Fante (Università degli studi di Ferrara, Italy)

Keywords: corpus-assisted discourse studies, forced lexical priming, delegitimization, evaluative cohesion

This paper is concerned with questions of evaluation in general and in particular with evaluative cohesion (Partington 2022; Partington forthcoming). Evaluation is fundamental to persuasion as the process of getting others to share one's evaluations. Evaluations can be built up in texts as part of a purposeful persuasive strategy and used to construct arguments (Hunston 2010). Hoey's lexical priming theory describes in detail (Hoey 2005) how we are primed by our encounters with language production, for the semantic associations, including evaluative associations, of lexis. Repeated evaluations foreground some aspects of a perceived reality, in such a way as to promote a particular definition of a problem, an interpretation, or a moral evaluation. When authors frequently repeat a certain form of evaluation to deliberately 'flood' the discourse with messages for a particular strategic purpose we can speak of forced priming (Duguid 2011; Duguid & Partington 2017; Duguid 2022), or an artificial boosting of the 'incremental effect of discourse' (Baker 2006: 13). The evaluation is frequently the point of the discourse and a cohesive factor in it. Although accumulations of evaluative lexis can be identified by quantitative methods (Labov 1972), such as sentiment analysis (Liu 2020), we still need close reading of the texts to be able to infer and characterise the strategies used to build an argument cohesively (Duguid & Partington 2024). In this case study we will use as an example the British Daily Mail, "Britain's most influential newspaper, the authentic voice of 'middle England'" (Henderson et al. 2016: 87)". The paper has been studied in the past for its role in the Brexit debate (Breeze 2018) and as a key combattant in what have been described by some as 'the culture wars' (Stoegner & Wodak 2015). The paper, like any news organ, can frame and reframe to suit its purpose and does so, formulaically, to construct its affiliations and delegitimise those constructed as enemies. It both echoes and grooms its readers. Here we examine salient items which emerge from applying a CADS approach to our dataset which comprises the 2021 section of the Daily Mail within the SiBol Corpus, accessible on SketchEngine (Kilgarriff et al. 2014) by adopting a corpus-assisted comparative perspective (Taylor & Marchi 2018; Partington, Duguid & Taylor 2013). The corpus size is 11 million tokens and 132,000 types. The delegitimising strategies of the Daily Mail thus emerge as a cohesively evaluative characteristic of the paper over many texts. We examine how these strategies, which include evaluative reporting verbs (whining, whinging, sneering, gloating, carping, clamouring, bleat crow, scoff), the attribution of affect (rage, fury, terror, horror) and the lexis of conflict (the animal rights brigade, the woke brigade, the politically correct brigade, woke class-war rabble, implacable, unappeasable) are employed to construct an imagined embattled position and how negative labelling (woke, smug, sanctimonious, virtue-signalling, hand-

wringing, pearl-clutching) is used to delegitimise opposing positions without rebuttal of the content of their argument.

References

- Baker, Paul. (2006). *Using corpora in discourse analysis*. London: Continuum.
- Breeze, Ruth. (2018). Enemies of the people: Populist performances in the Daily Mail reporting of the Article 50 case.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý and Vít Suchomel. (2014). The sketch engine: Ten years on. *Lexicography* 1: 7-36.
- Duguid, Alison. (2011). Insistent voices: Government messages corpus-assisted. In *Discourse Studies on the Iraq Conflict*. John Morley and Paul Bailey (eds.) Routledge, 250-276.
- Duguid, Alison. (2022). Counterpriming: changing evaluations through strategic forcing for persuasion. Paper given at symposium in honour of Michael Hoey. University of East Finland on 24th February 2022. <https://sites.uef.fi/pathwaystotextualitysymposium/symposium-papers-slides/>
- Duguid, Alison and Alan Partington. (2017). How they are produced and how they are received: Forced lexical primings in transdiscursive political messaging. In *Lexical priming: Applications and advances*. Michael Pace-Sigge and Katie J. Patterson (eds.) John Benjamins Publishing Company, 67-92. 10.1075/scl.79.03dug.
- Duguid, Alison and Alan Partington. (2023). New dogs, old tricks. A corpus-assisted study of the "art" of delegitimation in modern spoken political discourse. In *The Routledge handbook of discourse and disinformation*. Stefania Maci, Massimiliano Demata, Mark McGlashan and Philip Seargeant (eds.) Routledge. 10.4324/9781003224495
- Henderson, Ailsa, Jeffery, Charley, Robert Liñeira R, et al. (2016). England, Englishness and Brexit. *The Political Quarterly* 87(2): 187-199.
- Hoey, Michael. (2005). *Lexical priming. A new theory of words and language*. London: Routledge.
- Hunston, Susan. (2010). *Corpus approaches to evaluation: Phraseology and evaluative language*. New York: Routledge. [10.4324/9780203841686](https://doi.org/10.4324/9780203841686)
- Labov, William. (1973). *Sociolinguistic patterns*. Conduct and communication, v. 10. University of Pennsylvania Press. <https://books.google.it/books?id=hD0PNMu8CfQC>
- Liu, Bing. (2020). *Sentiment analysis: Mining opinions, sentiments and emotions*. 2nd ed. Cambridge: Cambridge University Press. 10.1017/9781108639286
- Partington, Alan, Alison Duguid and Charlotte Taylor. (2013). Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS). *Studies in Corpus Linguistics* 55. Amsterdam: John Benjamins Publishing Company.
- Taylor, Charlotte and Anna Marchi. (2018). *Corpus approaches to discourse: A critical review*. New York: Routledge.
- Partington, Alan. (forthcoming). Joined-Up' thinking and evaluation propositional and *evaluative* cohesion. In *The Language of Persuasion in Politics*. Alan Partington and Charlotte Taylor (eds.) Routledge.
- Partington, Alan. (2022). Evaluative cohesion' paper given at symposium in honour of Michael Hoey. University of East Finland 24 02 2022. <https://sites.uef.fi/pathwaystotextualitysymposium/symposium-papers-slides/>
- Stoegner, Karin and Ruth Wodak. (2016). "The man who hated Britain" – the discursive construction of "National Unity" in the Daily Mail'. *Critical Discourse Studies* 13(2): 193-209. [10.1080/17405904.2015.1103764](https://doi.org/10.1080/17405904.2015.1103764)

A useful methodological synergy?

Examining what makes a phrase legal, political or even altered by combining corpus linguistics with statutory interpretation

Remart Padua Dumlao (Monash University, Australia)

Keywords: statutory interpretation, legal text, modality

Although scholars have theorised corpus linguistic techniques as a new approach to legislative and constitutional interpretation (Solan 2020), researchers have not sufficiently investigated their methodology and actual application. My research explores the advantages of corpus linguistics as an alternative to legal interpretation and will attempt to fill this gap. The legality of this method arises from the widely accepted belief in courts that we should consider statutory terms in their ordinary sense when a statute does not expressly specify them or when definitions based on their usage in actual legal proceedings exist (Jack 2011; Bernstein 2021; Chomsky 2011). Corpus linguistics might offer a valuable technique for understanding the term's "new originalism" as well as its conventional meaning (Whittington 2004: 43), serving not only as a means to understand the terms themselves but also as a tool to comprehend how they circulate, transcend and translate in various communication forms.

In this paper, I focus on a specific incident in the Philippines, where a robust debate emerged regarding the meaning of the phrase "may be removed from office" in Section 2, Article XI of the 1987 Constitution, both before and during the filing of the impeachment complaint against Maria Lourdes Sereno (also known as Republic v. Sereno) in 2018. While these are not the sole grounds for her impeachment, they provide a starting point for discussing legal interpretation across various fields. I developed a specialized corpus named Philippine Legal Texts (PhiLeTex), which includes the constitution, laws and pronouncements and contains 203,210 tokens and I compared it to the PhilE corpus (Phil-Brown), comprising 674,000 words of Philippine non-legal texts (Collins, Borlongan & Yao 2014). This comparison enabled me to analyze the use and interpretation of the modal "may" in Philippine law. To identify what distinguishes the word 'may' from the most common modalities in Philippine English, I compared the PhiLetex to the Phil-Brown corpus (Collins, Borlongan & Yao 2014; Baker 2019). I then used the purpose of 'may' in legal texts as a criterion to determine whether the views of various Philippine legal professionals regarding the modal 'may' are driven by political hype, supported by legal justification, or based on shared understanding.

The study suggests that in the Philippine legal system, stakeholders contest and negotiate the interpretation of a legal phrase, like 'may,' often placing significant weight on both legal and shared understandings. This methodological and analytical approach should assist legal professionals and researchers interested in statutory interpretations by illuminating whether a particular word serves as a legal term or emerges from information manipulation or political scheming. I will also discuss some implications, suggestions for further research and valuable corpus linguistics techniques for analyzing legal texts.

The wolf ate us and the wolf being cancer

A methodological comparison of metaphor retrieval approaches

Nathan Dykes (Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany)

Keywords: metaphor, methodology, cancer, web corpus, semantic tagging

Metaphors have long been a source of interest in various areas of linguistics, including discourse analysis. Corpus-informed approaches play an important role in this context, since they can help to add nuance to established beliefs, such as the view that violence-based metaphors are always disempowering to patients (cf. Semino et al. 2015).

However, extracting a wide range of metaphors from a corpus is challenging: by definition, a study interested in retrieving a wide range of metaphor realisations from a certain domain will have to find a way to deal with the unpredictability of its lexical forms. A popular approach is to start from a corpus sample, manually identifying all metaphors of interest. The overall corpus is then annotated with the USAS tagger (Piao et al. 2008) and all tags featuring a relevant metaphor in the sample are checked on the entire dataset. USAS has been successfully applied to metaphor identification in various studies and genres (e.g. Demmen et al. 2015, Sun & Jan 2014). Methodologically speaking, Koller et al. (2008) have found it to achieve better recall than manual annotation of their overall data. However, recall of the USAS approach has not been quantified beyond this observation, i.e. there is no indication of how the proportion of relevant metaphors that one can expect to find this way. Moreover, USAS is currently the only established method to systematically search a thematic corpus for a large variety of target domain vocabulary (besides reading extensive amounts of corpus data). While other knowledge sources such as dictionaries are regularly used in corpus-based metaphor studies, they typically serve to verify the metaphor status of a candidate word rather than as a way of yielding additional candidates. An even more under-used resource are word embeddings, which can help to identify distributionally and thereby often semantically, similar lexis.

The study corpus consists of video transcriptions from RussiaToday's English YouTube channel, with all texts having cancer as their main topic. The automatically generated captions were extracted and converted to a corpus of ca. 250.000 tokens.

This study compares the extraction of metaphors with the USAS semantic tagger to two alternative methods. As a starting point for all three methods, the same random text sample is manually checked for violence metaphors.

The two alternative methods both aim to directly extend the list of metaphor candidates via external resources: in one case, via general-purpose dictionaries, in the other case via distributionally similar words according to word embeddings trained on large web corpora.

All three methods successfully retrieve a wide variety of both established and creative violence metaphors for cancer. At the same time, USAS is the only approach that hardly finds any target domain words which were not retrieved by either of the other

methods. As argued by Koller et al. (2008), semantic domains provide useful indicators as to whether a given tag is likely to yield additional target domain items that were not present in the manual sample. In terms of quantitative performance, the USAS approach is deemed to achieve the best compromise between recall and efficiency, while the dictionary method achieves the highest recall at ca. a third of all relevant metaphors.

References

- Demmen, Jane, Semino, Elena, Demjén, Zsófia, Koller, Veronika, Hardie andrew, Rayson, Paul and Sheila Payne. (2015). A computer-assisted study of the use of violence metaphors for cancer and end of life by patients, family carers and health professionals. *International Journal of Corpus Linguistics* 20(2): 205-231.
- Koller, Veronika, Hardie andrew, Rayson, Paul and Elena Semino. (2008). Using a semantic annotation tool for the analysis of metaphor in discourse. *Metaphorik. de* 15(1): 141-160.
- Piao, Scott, Rayson, Paul, Archer, Dawn, Wilson andrew and Tony McEnery. (2003). Extracting multiword expressions with a semantic tagger. In *Workshop on Multiword Expressions: Analysis, Acquisition and Treatment, at ACL 2003, 41st Annual Meeting of the Association for Computational Linguistics*: 49-56.
- Semino, Elena, Demjén, Zsófia and Jane Demmen. (2018). An integrated approach to metaphor and framing in cognition, discourse and practice, with an application to metaphors for cancer. *Applied linguistics* 39(5): 625-645.
- Sun, Ya and Jinlin Jiang. (2014). Metaphor use in Chinese and US corporate mission statements: A cognitive sociolinguistic analysis. *English for Specific Purposes* 33: 4-14.

Finding absence through double triangulation

Cross-linguistic corpus-assisted discourse analysis of the incest taboo in the French and British press (2017-2022)

Sophie Eyssette (Sapienza Università di Roma, Italy & Uniwersytet Śląski w Katowicach, Poland)

Keywords: corpus linguistics, critical discourse analysis, absence, cross-linguistics, incest taboo

60.000 children are victims of incest every year in France (CIIVISE, 2023). In the UK, no number is provided by the official authorities. The absence of data in the UK does not mean there is no incest; rather, it suggests that the prevalence of incest is still a strong taboo. Anthropologically, incest is either defined as marital prohibition (Durkheim 1987; Héritier

2012; Lévi-Strauss 1949), intrafamilial sexual abuse (Armstrong 1978; Dussy 2021; Gilgun 1995; Justice & Justice 1980; Russell 1983, 1999), or consensual intrafamilial relationship (Brey 2022; Ovidie 2022). Notably, linguistic taboos on the incest taboo have received limited attention in linguistics (Foubert, Marinato and Vallery 2022). This lack of research led to the investigation of the French and British press coverage on the incest taboo from 2017 to 2022. Furthermore, this paper aims to provide a methodology for finding absence in a corpus. It will be suggested to apply a double triangulation: a data triangulation between French and British corpora and a method triangulation between corpus linguistics and critical discourse analysis (see appendix).

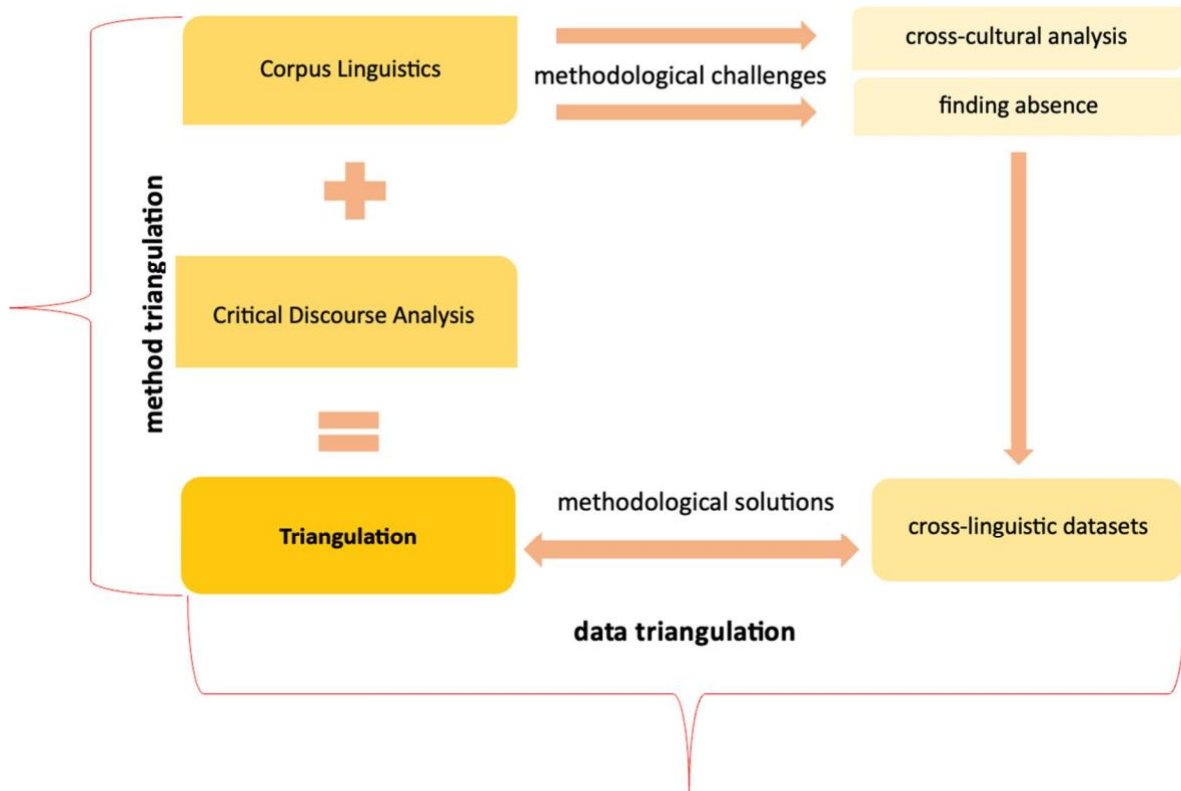
It has been widely discussed that the combination of corpus linguistics and critical discourse analysis offers several advantages for researchers, especially the ability to limit bias and yield more robust statistical findings (Baker 2023; Baker & Egbert 2016; Marchi & Taylor 2016; Vessey 2013). It is argued here that this triangulation allows researchers to find absence, as critical discourse analysis provides contextual elements to understand what is missing in a text (Strand 2018; Taylor 2014). Furthermore, some methods in corpus linguistics have been developed to investigate absence, one of which is comparison (Duguid & Partington 2018; Partington 2014). This paper aims to compare two different sets of language for the first time. As mentioned above, while statistics in France are available, they are missing in the UK, which leads us to question whether the mediatic coverage of the incest taboo is framed similarly or diversely.

The findings show that the British press tends to favour an erotic perception of incest, through the coverage of consensual intrafamilial relationships. However, in the French press, the emphasis is on child sexual abuse, due to a high-profile scandal in 2021 that changed the perception of the incest taboo abruptly. Therefore, the paper aims to show how this methodology of double triangulation provides double checks and balances to avoid overinterpreting absence in the corpus.

References

- Armstrong, Louise. (1978). *Kiss daddy goodnight: A speak-out on incest*. Hawthorn Books.
- Baker, Paul. (2023). *Using corpora in discourse analysis*. Bloomsbury Academic.
- Baker, Paul and Jesse Egbert. (2016). *Triangulating methodological approaches in corpus linguistic research*. Routledge.
- Brey, Iris. (2022). L'inceste qui crève les yeux. Représentations dans les séries et au cinéma. In *La Culture de l'inceste*. Éditions du Seuil.
- CIIVISE. 2023. *Violences sexuelles faites aux enfants: Le coût du déni*. Commission indépendante sur l'inceste et les violences sexuelles faites aux enfants.
- Duguid, Alison and Alan Partington. (2018). Absence, you don't know what you're missing. Or do you? In *Corpus Approaches to Discourse, A Critical Review*. Charlotte Taylor and Anna Marchi (eds.) Routledge.
- Durkheim, Émile. (1987). *La prohibition de l'inceste et ses origines: Étude de sociologie*. Éditions Payot & Rivages.
- Dussy, Dorothée. (2021). *Le berceau des dominations: Anthropologie de l'inceste*. Revised edition. Pocket.
- Foubert, Océane, Marinato, Lola and Robin Vallery. (2022). Liberation of speech or liberation of listening? #MeTooInceste testimonies and their reception on Twitter. Language, Ideology and Power research group seminar. Lancaster University.
- Gilgun, Jane F. (1995). We shared something special: The moral discourse of incest perpetrators. *Journal of Marriage and the Family* 57(2): 265. 10.2307/353682
- Héritier, Françoise. (2012). *Les deux sœurs et leur mère: Anthropologie de l'inceste*. O. Jacob.
- Justice, Blair and Rita Justice. (1980). *The broken taboo, sex in the family*.
- Lévi-Strauss, Claude. (1949). *Nature, culture et société: Les structures élémentaires de la parenté*, chapitres I et II. Presses Universitaires de France.
- Marchi, Anna and Charlotte Taylor. (2016). If on a winter's night two researchers... A challenge to assumptions of soundness of interpretation. *Critical Approaches to Discourse Analysis Across Disciplines* 67: 372-373. 10.5209/CLAC.53494
- Ovidie. 2022. Petite sémiotique de la step mom. Pornographies et culture de l'inceste. In *La Culture de l'inceste*. Éditions du Seuil.
- Partington, Alan. (2014). Mind the gaps: The role of corpus linguistics in researching absences. *International Journal of Corpus Linguistics* 19(1): 118-146. 10.1075/ijcl.19.1.05par
- Russell, Diana E. H. (1983). The incidence and prevalence of intrafamilial and extrafamilial sexual abuse of female children. *Child Abuse & Neglect* 7(2): 133-146. 10.1016/0145-2134(83)90065-0
- Russell, Diana E. H. (1999). *The secret trauma: Incest in the lives of girls and women; with a new introduction by the author*. Revised edition. Basic Books.
- Strand, Cecilia. (2018). Cross-media studies as a method to uncover patterns of silence and linguistic discrimination of sexual minorities in Ugandan print media. In *Exploring Silence and Absence in Discourse*. Melani Schröter and Charlotte Taylor (eds.) Palgrave, 125-157.
- Taylor, Charlotte. (2014). Investigating the representation of migrants in the UK and Italian press: A cross-linguistic corpus-assisted discourse analysis. *International Journal of Corpus Linguistics* 19(3): 368-400. 10.1075/ijcl.19.3.03tay
- Vessey, Rachelle. (2013). Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora* 8(1): 1-26. 10.3366/cor.2013.0032

Appendix



Schema of the double triangulation for finding absence in corpora

The discourse of inclusion & disability on social media on the example of a German corpus of 214,926 Tweets on #disability and #inclusion between 2007-2023

Annamária Fábíán (Universität Bayreuth, Germany)

Igor Trost (Universität Passau, Germany)

Keywords: #DisabilityDiscourse, #DiscourseofInclusion, #SentimentAnalysis, #DigitalDiscourseAnalysis, #CorpusLinguistics

Whereas there is a wealth of studies on the language of discrimination, particularly within discourse studies, research addressing specifically the linguistic strategies of individuals belonging to different diversity dimensions for inclusion are comparatively rare. At the same time, many linguistic studies make use of data from digital media, itself an increasingly popular object of study in linguistics (Crystal 2011; Zappavigna 2012; Marx & Weidacher 2014; Bubenhofer 2017; De Decker & Vandekerckhove 2017; Jang et al. 2019; Wang & Taylor 2019; Abel & Glaznieks 2020; Marx et al. 2020; Wright 2020). On the one hand, studies on linguistic strategies of diverse individuals for inclusion are continuously rising, on the other hand digitalization provides more visibility to diversity as ever before, which faces academia to keep up with the emergent need of more linguistic insights into the rapidly developing linguistic strategies of diverse individuals for inclusion on social media. This study overcomes this emergent issue by introducing a corpus linguistic examination of these digital linguistic strategies with qualitative and quantitative (corpus-driven) methods and examines the digital discourse of inclusion and disability on social media. For this corpus linguistic analysis, we use a corpus of 214,926 Tweets published in the time span of 2007-2023 under the hashtags `inclusion` and `disability`. This unique study in applied (German) linguistics provides valuable insights into the vocabulary of inclusion and disability and the co-occurrences of the lexical units and the n-grams by using AntConc and CorpusExplorer. In addition, our research also includes emotions in the German digital discourse on disability and inclusion and shows a classification of the Tweets via Sentiment Analysis (SentiStrength). The presentation highlights the relevance of a quantitative and qualitative examination of the (German) social media communication of inclusion and disability and for this, also the use of tools eligible for the lexical and sentiment analysis of big language data from social media on the example of the German digital discourse on inclusion and disability on Twitter/X in the time period of appr. 15 years. This urgent scientific contribution also aims at a significant methodological value to representative discourse studies as discourse linguistic interest in emotions has rapidly started raising but the knowledge of particular software-driven methods are still limited. In addition, we present also the challenges regarding classificatory issues of the vocabulary when using Software for the representative analysis of emotions and however, methodological solutions for more valid representative lexical analysis in digital discourses.

References

- Abel andrea and Aivars Glaznieks. (2020). Textqualität in Sozialen Medien. In *Deutsch in Sozialen Medien: Interaktiv – multimodal – vielfältig*. Konstanze Marx, Henning Lobin and Axel Schmidt (eds.) Berlin, Boston: De Gruyter, 53-74.
- Bubenhofer, Noah. (2017). Kollokationen, n-Gramme, Mehrworteinheiten. *Handbuch Sprache in Politik und Gesellschaft* 19.
- Crystal, David (2011). *Internet linguistics. A student guide*. London: Routledge.
- De Decker, Benny and Reinhild Vandekerckhove. (2017). Global features of online communication in local Flemish: Social and medium-related determinants. *Folia linguistica* 51(1): 253-281.
- Jang, S Mo, Brooke W McKeever, Robert McKeever and Joon Kyoung Kim. (2019). From social media to mainstream news: The information flow of the vaccine-autism controversy in the US, Canada and the UK. *Health communication* 34(1): 110-117.
- Marx, Konstanze and Georg Weidacher. (2014). *Internetlinguistik. Ein Lehr- und Arbeitsbuch*. Tübingen: Günter Narr.
- Marx, Konstanze, Henning Lobin and Axel Schmidt (2020). *Deutsch in Sozialen Medien: interaktiv – multimodal – vielfältig*. Berlin, Boston: De Gruyter.
- Wang, Jiayi and Charlotte Taylor. (2019). The conventionalisation of mock politeness in Chinese and British online forums. *Journal of Pragmatics* 142: 270-280.
- Wright, David (2020). The discursive construction of resistance to sex in an online community. *Discourse, Context & Media* 36: 100402.
- Zappavigna, Michele (2012). *Discourse of Twitter and social media: How we use language to create affiliation on the web* 6. A&C Black.

A first step towards a better planet

Phraseological patterns of anticipatory discourse in youth environmentalism

Mariasophia Falcone (Università degli Studi di Modena e Reggio Emilia, Italy)

Keywords: anticipatory discourse, phraseology, GenZ, agency

The recent global environmental protests, sparked by Greta Thunberg's Friday school strikes in 2018, have significantly mobilised younger generations, notably Generation Z (GenZ), in support of the environmental cause (Wahlström et al. 2019). Consequently, youth environmentalism has surged globally, with members of GenZ growing increasingly concerned with environmental themes. As younger generations clearly have a long-term stake in environmental issues and will bear the brunt of current environmental decisions, an increase in eco-anxiety, along with a growing worry about the future have been noticed as some of the outcomes of this rise in environmental awareness (Seemiller & Grace 2018; Clayton 2020; Clayton & Karazsia 2020). In addition, future-oriented issues are particularly central to environmentalism itself, given that, unlike other political struggles, it is uniquely centred on securing the future and safeguarding its very existence to allow life on earth (O'Brien & Lousley 2017). In light of this, investigations into the discursive construction of the future (anticipatory discourse), defined as texts or utterances oriented towards future events, actions, or states (de Saint-Georges 2012), within GenZ environmental activism appear quite promising. Therefore, the present study investigates recurring phraseology of anticipatory discourse in a corpus of blog posts of three GenZ environmental organizations published between 2018 and 2022 (GCWebC). The analysis follows a corpus-driven approach (Tognini-Bonelli 2001) and adopts a closed-class keywords only methodology for discourse analysis (Groom 2010). More specifically, grammatical keywords have been used as a starting point into the analysis of phraseological patterns which have revealed a key presence of future-related phraseology. Furthermore, as the analysis of anticipatory discourse involves considering the stance conveyed within the text in relation to the degree of agency supposed, leaning more towards agentive or fatalistic (de Saint-Georges 2012), attention has also been directed towards interpreting the organizations' agency towards future events or consequences, mainly in terms of their capability to take proactive measures in forthcoming situations or influence specific outcomes. Eventually, findings have revealed, on the one hand, a prevalence of phraseological patterns with epistemic stance portraying the future more as a consequence than as a possibility. Meanwhile, on the other, a somewhat ambivalent agency has been noticed: in terms of individual or collective behaviour, especially towards institutions, GenZ environmental organizations adopt a more agentive-leaning approach, while leaning towards rather fatalistic in relation to the state of the Earth and the ecosystem.

References

- Clayton, Susan and Bryan T. Karazsia. (2020). Development and validation of a measure of climate change anxiety. *Journal of Environmental Psychology* 69. [10.1016/j.jenvp.2020.101434](https://doi.org/10.1016/j.jenvp.2020.101434)
- Clayton, Susan. (2020). Climate anxiety: Psychological responses to climate change. *Journal of Anxiety Disorders* 74. [10.1016/j.janxdis.2020.102263](https://doi.org/10.1016/j.janxdis.2020.102263)
- Groom, Nicholas. (2010). Closed-class keywords and corpus-driven discourse analysis. In *Studies in corpus linguistics* 41. Marina Bondi and Mike Scott (eds.) Amsterdam: John Benjamins Publishing Company, 59-78. [10.1075/scl.41.05gro](https://doi.org/10.1075/scl.41.05gro)
- O'Brien, Susie and Cheryl Lousley. (2017). A history of environmental futurity: Special issue introduction. *Resilience: A Journal of the Environmental Humanities* 4(2-3): 1. [10.5250/resilience.4.2-3.0001](https://doi.org/10.5250/resilience.4.2-3.0001)
- Saint-Georges, Ingrid De. (2012). Anticipatory discourse. In *The Encyclopedia of Applied Linguistics*. Carol A. Chapelle (eds.) Wiley. [10.1002/9781405198431.wbeal0032](https://doi.org/10.1002/9781405198431.wbeal0032)
- Seemiller, Corey and Meghan Grace. (2019). *Generation Z: A century in the making*. Milton Park, Abingdon, Oxon: Routledge, Taylor and Francis Group.
- Tognini-Bonelli, Elena. (2001). *Corpus linguistics at work*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Wahlström, Mattias, Sommer, Moritz, Kocyba, Piotr, De Vydt, Michiel, de Moor, Joost, Davies, Stephen, Wouters, Ruud, Wennerhag, Magnus, Stekelenburg, Jacquélien, Uba, Katrin, Saunders, Clare, Rucht, Dieter, Mikecz, Daniel, Zamponi, Lorenzo, Lorenzini, Jasmine, Kołczyńska, Marta, Haunss, Sebastian, Giugni, Marco, Gaidyte, Teodora and Aron Buzogany. (2019). Protest for a future: Composition, mobilization and motives of the participants in Fridays for Future climate protests on 15 March, 2019 in 13 European cities.

Frames and framings of German extremism discourse

A distributional-semantic method for the corpus-driven exploration of semantic frames

Tim Feldmüller (Leibniz-Institut für Deutsche Sprache, Germany)

Keywords: extremism, corpus-driven, frame-semantics, word embeddings, collocation networks

The contribution presents results of a research project that has developed a corpus-driven (Tognini-Bonelli 2001) method to reconstruct frames and framings of the German extremism discourse of the period 1999 – 2021. Based on the distributional hypothesis "that there is a correlation between distributional similarity and meaning similarity" (Sahlgren 2008: 33) and Tognini-Bonelli's claim "that the choice patterns of words in text can create new, large and complex units of meaning" (2001:102), an attempt was made to interpret distributional patterns as frame-semantic structures in the sense of Busse (2012) - a complex frame model that represents a synthesis of, among others, Fillmore's, Minsky's and Barsalous' frame designs. A combination of methods aimed at the syntagmatic distribution of words (collocations) with those that systematically analyze their paradigmatic distribution (word embeddings) has proved to be particularly fruitful.

Word embeddings (Mikolov et al. 2013) are a way of reversing a collocation analysis. Instead of asking which words co-occur with a given word, they provide information about which words fit into similar contexts. Technically, they represent n-dimensional vectors that are embedded in a common vector space. Proximity in this vector space indicates a similar collocation profile and thus semantic equivalence (Bubenhof 2020). By applying a clustering algorithm (*k-Means*), neighborhoods can be identified in the vector space. In this way, word clusters that can be interpreted as frame elements (FE) of an extremism frame can be determined, for example actors (Terrormiliz | Terrormiliz_iS | Terrormiliz_islamischer_Staat / *terror_militia* | *terror_militia_ISIS* | *terror_militia_islamic_state*), actions (Attentat | Terroranschlag | Anschläge / *attack* | *terror_attack* | *attacks*) or characteristics (Antisemitismus | Rassismus | Rechtsextremismus / *antisemitism* | *racism* | *right-wing_extremism*) of extremism (variants).¹ At the same time, the contained words represent lexical units that evoke the frame.

Collocations in turn make it possible to analyze the relationship between the FEs. To analyze not only isolated syntagmatic relationships but "a complex network of semantic relationships" (Brezina, McEnery & Wattam 2015: 141), the mutual collocations of the clusters were calculated and converted into a collocation network. The nodes of the network represent the WE clusters, while the edges are based on their collocation values

¹ Clusters labeled after their three most central words in the vector space. Statistically significant bi- and trigrams have been merged before the training.

(log-likelihood). The network can also be analyzed using network analytic methods, as advocated by Baker (2016); here, an automatic community detection algorithm (*spin glass*) has proved fruitful in identifying thematic sub-frames that can inform framing analysis.

In order to investigate the framing of extremism variants, which has been shown to be extremely variable (Ackermann et al. 2015), a large corpus (1.3 billion tokens) was constructed from three German newspapers (*Taz*, *Spiegel* and *Welt*) for the period 1999 – 2021 and both diachronic and newspaperspecific sub-corpora were built. One example for changes in framing on the diachronic axis, is that the potential for violence of right-wing extremism after 9/11 is lost from the media spotlight for many years - a finding that is already evident in the macrostructure of the frame and is congruent with existing discourseanalytical literature (Czulo, Nyhuis & Weyell 2020).

References

- Ackermann, Jan, Behne, Katharina, Buchta, Felix, Drobot, Marc and Philipp Knopp. (2015). *Metamorphosen Des Extremismusbegriffes: Diskursanalytische Untersuchungen Zur Dynamik Einer Funktionalen Unzulänglichkeit*. Wiesbaden: Springer VS.
- Baker, Paul. (2016). The shapes of collocation. *International journal of corpus linguistics* 21(2): 139-164. 10.1075/ijcl.21.2.01bak
- Brezina, Vaclav, McEnery, Tony and Stephen Wattam. (2015). Collocations in context: A new perspective on collocation networks'. *International journal of corpus linguistics*: 139-173. 10.1075/ijcl.20.2.01bre
- Bubenkofer, Noah. (2020). Semantische Äquivalenz in Geburtserzählungen: Anwendung von Word Embeddings'. *Zeitschrift Für Germanistische Linguistik* 48(3): 562-589. 10.1515/zgl2020-2014
- Busse, Dietrich. (2012). *Frame-Semantik: Ein Kompendium*. Berlin: De Gruyter.
- Czulo, Oliver, Nyhuis, Dominic and Adam Weyell. (2020). Der Einfluss extremistischer Gewaltereignisse auf das Framing von Extremisten auf SPIEGEL Online. *Journal für Medienlinguistik* 3(1): 14-45. 10.21248/JFML.2020.11.2
- Mikolov, Tomas, Chen, Kai, Corrado, Greg and Jeffrey Dean. (2013). Efficient estimation of Word representations in Vector Space'. arXiv. 10.48550/ARXIV.1301.3781
- Sahlgren, Magnus. (2008). The distributional hypothesis'. *Rivista Die Linguistica* 20(1): 33-53.
- Tognini-Bonelli, Elena. (2001). *Corpus linguistics at work. Studies in Corpus Linguistics* 6. Amsterdam: John Benjamins. 10.1075/scl.6

How humans vs. machines identify discourse topics

A methodological triangulation

Mathew Gillings (Wirtschaftsuniversität Wien, Austria)

Sylvia Jaworska (University of Reading, United Kingdom)

Keywords: concordance analysis, topic modelling, ChatGPT, close reading, triangulation

Identifying discourses and discursive topics in a set of texts has not only been of interest to linguists, but to researchers working across the full breadth of the social sciences. Traditionally, these analyses have been conducted based on small-scale interpretive analyses of discourse which involve some form of close reading. Naturally, however, that close reading is only possible when the dataset is small and it leaves the analyst open to accusations of bias, cherry-picking and a lack of representativeness (Mautner 2015).

Designed to avoid these issues, other methods have emerged which involve larger datasets and have some form of quantitative component. Within linguistics, this has typically been through the use of corpus-assisted methods, whilst outside of linguistics, topic modelling is one of the most widely-used approaches. Increasingly, researchers are also exploring the utility of LLMs (such as ChatGPT) to assist analyses (Curry et al. 2023). How corpus linguistics, topic modelling and LLM-assisted work differ, though, is in the degree of contextualisation available to the researcher. Topic modelling algorithms reduce texts to a simple bag-of-words, presenting only a list of co-occurring words to the researcher for analysis. Researchers utilising topic modelling typically eyeball these words and attempt to ascertain topic labels (Gillings & Hardie 2022). On the other hand, corpus-assisted methods and in particular concordance analysis, allow the user to see words of interest within their co-text (typically a few words on either side). Corpus-assisted methods, then, are somewhere in between the completely decontextualised topic modelling and the completely contextualised close reading.

This talk reports on a study assessing the effect that analytical method has on the interpretation of texts, specifically in relation to the identification of the main topics. Using a corpus of corporate sustainability reports, totalling 98,277 words, we asked 6 different researchers, along with ChatGPT, to interrogate the corpus and decide on its main 'topics' via four different methods. Each method gradually increases in the amount of context available.

- Method A: ChatGPT is used to categorise the topic model output and assign topic labels;
- Method B: Two researchers were asked to view a topic model output and assign topic labels based purely on eyeballing the co-occurring words;
- Method C: Two researchers were asked to assign topic labels based on a concordance analysis of 100 randomised lines of each co-occurring word;
- Method D: Two researchers were asked to reverse-engineer a topic model output by creating topic labels based on a close reading.

The talk explores how the identified topics differed both between researchers in the same condition and between researchers in different conditions shedding light on some of the mechanisms underlying topic identification by machines vs. humans or machines assisted by humans. Ultimately, we find that the more context is available, the more divergent the interpretations of the text. We conclude with a series of tentative observations regarding the benefits and limitations of each method and recommendations for researchers when it comes to choosing an analytical technique for the identification of discourse topics.

References

- Curry, Niall, Baker, Paul and Gavin Brookes. (2023). Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics*.
- Gillings, Mathew and Andrew Hardie. (2022). The interpretation of topic models for scholarly analysis: An evaluation and critique of current practice. *Digital Scholarship in the Humanities*. 10.1093/lc/fqac075
- Mautner, Gerline. (2015). Checks and balances: How corpus linguistics can contribute to CDA. In *Methods of Critical Discourse Analysis*. Ruth Wodak and Michael Meyer (eds.) London: Sage, 154-179.

Corpus-assisted framing analysis of Italian newspaper discourse on diabetes before and during COVID-19 pandemic

A case study on La Repubblica

Luca Giordano (Università degli Studi di Napoli L'Orientale, Italy)

Keywords: framing, Italian, newspapers, diabetes, COVID-19

The term “diabetes” encompasses type 1, gestational and type 2 diabetes, the most common. All diabetic subjects, regardless of type, risk serious COVID-19 complications¹. In the news, the important type distinction is often ignored and the causes for the condition are framed as mostly behavioural rather than also medical and societal, promoting blame.

This study answers two research questions (RQs):

- RQ1: Does the relative frequency of diabetes types in La Repubblica’s coverage before and during COVID-19 pandemic reflect type 2’s prevalence among diabetic patients?
- RQ2: Was diabetes framed any differently before and during COVID-19 pandemic by Italian newspaper La Repubblica?

The main contributions of this study are 1) to fill the research gap on this topic in the Italian context to allow comparisons with other findings and 2) to shed light on the framing of diabetes in newspaper discourse during the COVID-19 pandemic.

The data, processed with AntConc², is a subset of the TRADISAN corpus² (Giordano & di Buono 2023), a collection of 32,101 health-related Italian news published by 38 newspapers. From La Repubblica subset two timeframe-defined subcorpora are created, containing respectively articles published within the 3 years before the pandemic, 2017-2018-2019 and during the 3 years of pandemic in Italy, 2020-2021-2022 (Table 1). The reference corpus is ItWac (Baroni et al. 2009).

Both quantitative and qualitative analyses are conducted. In the quantitative analysis, first a keyword list is generated for each subcorpus to observe the rank and keyness by likelihood of the term “diabete” (Table 2). Next, the raw frequencies in each subcorpus of the umbrella term “diabete” with all its variants are computed (Table 3). Finally, the collocational profiles of the term “diabete” in the subcorpora are analysed to highlight some interesting linguistic patterns (Table 4).

In the qualitative analysis, first the rankings and keyness of the term in the two keyword lists are compared. Next, raw frequency differences and the importance of diabetes type distinction are discussed to answer RQ1; follow examples and finally the framing analysis is conducted to answer RQ2.

¹ [American Diabetes Association - How COVID-19 Impacts People with Diabetes](#) ² [AntConc](#)

² [TRADISAN corpus](#)

The results of the quantitative and qualitative analyses show that:

- Diabetes coverage (keyness) by La Repubblica decreased during the COVID-19 pandemic;
- The type distinction was almost always ignored. When made, type 2 was the most frequently mentioned (answering RQ1), followed by type 1. Gestational diabetes was never mentioned;
- Regarding RQ2, before the pandemic two frames were mainly used: *Behavioural* and *Fight & Control*. Diabetes was framed as an individual's battle: examples show that diabetes "attacks", "generates risk" and the patient "controls", "protects" and "suffers". This puts heavy emphasis on individual behaviour and might perpetrate blame (Broom & Whittaker 2004; Gollust & Lantz 2009; Stefanik-Sidener 2013; Gounder & Ameer 2018);
- During the pandemic, two different frames were mainly used: *Medical* and *Related pathologies*. The focus shifted to related pathologies, with high reference to COVID-19 infection and heart-related complications. Heavy emphasis was put to the correlation between diabetes and COVID-19, highlighting purely medical factors instead of only behavioural ones.

References

- Baroni, Marco, Bernardini, Siliva, Ferraresi, Adriano and Eros Zanchetta. (2009). The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation* 43(3): 209-226. [10.1007/s10579-009-9081-4](https://doi.org/10.1007/s10579-009-9081-4)
- Broom, Dorothy and Andrea Whittaker. (2004). Controlling diabetes, controlling diabetics: Moral language in the management of diabetes type 2. *Social Science & Medicine* 58(11): 2371-2382. [10.1016/j.socscimed.2003.09.002](https://doi.org/10.1016/j.socscimed.2003.09.002)
- Giordano, Luca and Maria Pia D. Buono. (2023). Assessing Italian news reliability in the health domain through text analysis of headlines. Proceedings of the 1st International Workshop on Disinformation and Toxic Content Analysis (DiTox 2023). *Language, Data and Knowledge*.
- Gollust, Sarah E. and Paula M. Lantz. (2009). Communicating population health: Print news media coverage of type 2 diabetes. *Social Science & Medicine* 69(7): 1091-1098. [10.1016/j.socscimed.2009.07.009](https://doi.org/10.1016/j.socscimed.2009.07.009)
- Gounder, Farzana and Rashid Ameer. (2018). Defining diabetes and assigning responsibility: How print media frame diabetes in New Zealand. *Journal of Applied Communication Research* 46(1): 93-112. [10.1080/00909882.2017.1409907](https://doi.org/10.1080/00909882.2017.1409907)
- Stefanik-Sidener, Kelsey. (2013). Nature, nurture, or that fast food hamburger: Media framing of diabetes in the New York Times from 2000 to 2010. *Health Communication* 28(4): 351-358. [10.1080/10410236.2012.688187](https://doi.org/10.1080/10410236.2012.688187)

Appendix

Subcorpus	# Articles	# Tokens
La Repubblica pre-covid (2017/18/19)	311	41270
La Repubblica during-covid (2020/21/22)	2548	880864

Table 1 – Subcorpora

Subcorpus	Keyword list rank	Keyness (Likelihood)
La Repubblica pre-covid (2017/18/19)	#13/418	183.733
La Repubblica during-covid (2020/21/22)	#318/1603	129.150

Table 2 – “Diabete” keyword in the data

Query	La Repubblica pre-covid (2017/18/19)	La Repubblica during-covid (2020/21/22)
diabete	35	121
diabete di tipo 1	0	1
diabete di tipo I	0	0
diabete infantile	0	0
diabete giovanile	0	1
diabete insulino-dipendente	0	2
diabete mellito di tipo 1	0	0
diabete mellito di tipo I	0	0
diabete tipo 1	0	0
diabete tipo I	0	0
diabete di tipo 2	3	19
diabete di tipo II	0	0
diabete mellito di tipo 2	1	0
diabete mellito di tipo II	0	0
diabete tipo 2	0	2
diabete tipo II	0	0
diabete gestazionale	0	0
TOTAL	39 (range 21)	146 (range 70)

Table 3 – Raw frequencies of “diabete” and variants

			Precovid					
Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect	
il	1	29	21	8	14	38.156	1.993	
cuore	2	9	5	4	6	30.722	3.777	
2	3	4	0	4	4	18.959	4.786	
proteggi	4	2	1	1	1	18.284	7.873	
controllare	5	3	2	1	2	16.970	5.458	
soffre	5	3	3	0	3	16.970	5.458	
rischio	7	7	4	3	5	15.823	2.861	
attacca	8	2	0	2	1	15.306	6.873	
cardiovascolari	9	3	1	2	3	14.567	4.873	
tipo	10	4	0	4	4	14.108	3.873	
			Duringcovid					
Collocate	Rank	FreqLR	FreqL	FreqR	Range	Likelihood	Effect	
ipertensione	1	18	8	10	17	145.119	7.222	
tipo	2	24	1	23	17	129.258	5.274	
2	3	24	1	23	17	87.179	3.958	
cardiovascolari	4	12	2	10	10	78.779	6.144	
obesità	5	13	9	4	12	78.554	5.762	
malattie	6	16	7	9	14	63.772	4.232	
il	7	66	50	16	35	43.169	1.326	
patologie	8	9	5	4	7	37.007	4.331	
con	9	37	30	7	20	32.534	1.588	
ipercolesterolemia	10	3	2	1	3	28.698	8.300	
croniche	11	5	3	2	5	28.633	5.535	
sovrappeso	12	5	3	2	5	27.804	5.413	
cuore	13	7	4	3	4	27.800	4.225	
persone	14	13	13	0	7	27.660	2.759	
come	15	22	20	2	20	27.199	1.956	
rischio	16	15	8	7	12	25.691	2.399	
annoverare	17	2	1	1	1	24.994	10.300	
proteggi	17	2	1	1	1	24.994	10.300	
tipo2	19	2	0	2	2	22.002	9.300	
cardiache	20	3	1	2	3	20.841	6.426	
insulino	21	2	0	2	1	20.313	8.715	
iperglicemia	22	2	1	1	1	18.220	7.979	
associate	23	3	3	0	3	17.753	5.676	
soffrono	24	3	3	0	3	17.595	5.637	

giovanile	25	2	0	2	1	17.481	7.715
scompenso	26	3	2	1	2	16.599	5.394
controllare	27	3	2	1	2	15.751	5.185
cardiaco	28	3	2	1	2	15.215	5.053

Table 4 – Collocational profiles of “diabete” in pre- and during-covid subcorpora

The contextual variables of academic spoken discourse

A metadiscursive perspective

Tuğçe Aslıhan Güler (Middle East Technical University, Turkey)

Çiler Hatipoğlu (Middle East Technical University, Turkey)

Keywords: metadiscourse, academic spoken discourse, corpus-assisted discourse analysis

This study depicts a detailed picture of metadiscourse functions of speech that participants employ during dialogic speech events in a specific academic discourse. With this aim, the focus of the study is to extensively analyze discourse functions of metadiscourse that participants use in formal meetings and to examine to what extent the contextual variables of academic discourse such as participant roles, type of interaction and participants' alignment to or distance from each other affect their use of metadiscourse.

To analyze the discourse functions and explore the dynamics inherent in academic settings, a spoken corpus of thesis advisory meetings has been compiled for this mixed-method study. The data, a corpus of 23.667 words (176 minutes), collected from two advisory meetings through video recordings and were transcribed and analyzed with MAXQDA. A total of count of 1655 instances of code frequencies revealed through the qualitative analysis. The dataset was categorized by gender, the role of participants, use of inclusive language and discourse functions of metadiscourse (Adel 2023).

The findings revealed diverse patterns concerning 'we' references made by participants. Regarding randomly coded segments of 'inclusive use of language' through pronoun 'we', while candidates and supervisors predominantly employ inclusive 'we' when referring to each other, demonstrating a discourse reflexivity (Mauranen 2010) and co-operation (Fortanet 2004), internal and external committee members tend to address the entire committee using exclusive 'we', as a distancing tool between the committee members and candidates. Metadiscourse functions were then systematically coded using Adel's taxonomy (2023). Within the discourse functions encompassed by the domain of 'words and their meanings', the predominant function identified was *clarifying*, evidencing a variety with monologic speech. Subsequent to this, an analysis was conducted to investigate discourse functions associated with topics and information flow. Within this subcategory, all five discourse functions were coded with their additional subgroups for *managing topic* and *managing phorics*. A notable amount of coded discourse functions in the dataset fall under managing phorics. *Comments on discourse* constituted an additional category within the category of topics and information flow. Notably, the most recurrent discourse function coded was *labelling speech acts*, followed by *contextualizing the discourse* featuring dominantly. The audience orientation category predominantly involved *acknowledging the audience*, in line with the previous findings on monologic speech events (Adel 2023). This prevalence is largely attributable to the dialogic nature of speech, evident in the frequency of codes such as 'thank you' or 'congratulations'.

Remarkably, *managing audience behavior* constituted the most frequently occurring discourse function. Concluding the taxonomy, the discourse functions of labelling author persona required further exploration, given the dynamic and adaptable nature of dialogic speech, allowing for the introduction of new functions and supplementary assumptions. The findings were contextualized within the academic spoken discourse, providing a thorough and nuanced understanding of how metadiscourse is employed, the diverse range of functions it serves in dialogues and conversations and the effects of contextual variables on the use of metadiscourse.

References

- Annelie, Adel. (2023). Adopting a 'move' rather than a marker approach to metadiscourse: A taxonomy for spoken student presentations. *English for Specific Purposes* 69: 4-18. 10.1016/j.esp.2022.09.001
- Fortanet, Inmaculada. (2004). The use of 'we' in university lectures: Reference and function. *English for Specific Purposes* 23: 45-66. 10.1016/S0889-4906(03)00018-8
- Mauranen, Anna. (2010). Discourse reflexivity – A discourse universal? The case of ELF. *Nordic Journal of English Studies* 9: 13-40. 10.35360/njes.216

Constructing discursive opposition

A case study of affixal negation in Ta-Nehisi Coates's discourse on race

Yulia Hathaway (Universitetet i Bergen, Norway)

Keywords: affixal negation, corpus-assisted discourse study, U.S. discourse on race, Ta-Nehisi Coates

This paper presents a corpus-based study of the use of the most productive negative affixes in English (*un-*, *non-* and *anti-*) - in a particular domain: discourse on race in the USA. It also discusses methodological considerations for corpus investigations of affixal negation.

Negation is conceptualized as a phenomenon of semantic opposition (Horn and Wansing 2020). However, negation and its subtypes have rarely been investigated in discourse studies and affixal negation has been an understudied subtype of linguistic negation (Joshi 2020). Furthermore, alternative discourses (e.g. antiracist discourse) have not been widely analyzed in (critical) discourse studies (Van Dijk 2023). Nevertheless, some scholars argue that the study of negation can offer insights into production and reproduction of ideologies (Nahajec 2021). This paper describes linguistic properties of affixal negation in a specialized corpus and discusses how this type of negation is used in an oppositional type of discourse to challenge ideological and societal norms.

Alternative discourses are often fragmented requiring a compilation of specialized corpora. This study analyzes the works of Ta-Nehisi Coates, a contemporary African American writer, recognized as “the most important essayist in a generation [...] who changed the national political conversation about race” (Fear 2019). *The Corpus of the non-fictional writings by Ta-Nehisi Coates* (COCO) is comprised of 349 full-text online articles from 10 U.S. news outlets and one monograph (468,899 words), covering a period of 22 years (1996-2018). Methodologically, the study combines corpus linguistic techniques with contextual(ized) discourse analysis. However, prefixed words are not easily extracted from large datasets. The challenge of identifying relevant tokens is further complicated by typographical variation in negative prefixation in English (with/without a hyphen or space) and infrequency of tokens (Duguid 2010). Therefore, a two-step search procedure, using *WordSmith tools 7.0*, was employed to identify all the relevant tokens of negative affixation in COCO. Based on the notions of lexicalization and productivity, the prefixed forms with *un-* (898 tokens), *non-* (122) and *anti-* (95) were selected for contextual analysis.

Results show that, in COCO, the three prefixes attach relatively freely to a variety of bases:

individual bases (e.g. *nonevent*), compounds (e.g. *anti-crime wave*) and phrases (e.g. *un-first-ladylike*), indicating that affixal negation can be used to produce novel or low-frequency formations in appropriate contexts. Such lexical creativity contributes to “the construction of discourse meaning... in subtle and interesting ways” (Munat 2007: xiv). Furthermore, even though the terms *racism* (188) and *racist* (111) occur throughout the

corpus (1997-2018), their prefixed antonyms are not attested prior to 2013. The term *non(-)racist* (5) has more negative connotations as it subtly indicates denial of or a passive attitude toward racism; whereas the use of *anti(-)racism* (4) and *anti(-)racist* (11) indicates a call for action against racism, which points to Coates's antiracist discourse as a part of a social movement from 2013 onwards.

References

- Duguid, Alison. (2010). Investigating anti and some reflections on modern diachronic corpus-assisted discourse studies (MD-CADS). *Corpora* 5(2): 191-220. [10.3366/cor.2010.0105](https://doi.org/10.3366/cor.2010.0105)
- Fear, David. (2019). 'The water dancer': Ta-Nehisi Coates' American odyssey. *Rolling Stone*. <https://www.rollingstone.com/culture/culture-features/tanehisi-coates-novel-water-dancerbook-review-889295/> (Accessed April 18, 2024)
- Horn, Laurence R. and Heinrich Wansing. (2020). Negation. [webpage]. *The Stanford Encyclopedia of Philosophy*. Stanford University. <https://plato.stanford.edu/archives/spr2020/entries/negation/> (Accessed January 8, 2024).
- Joshi, Shrikant. (2020). Affixal negation. In *The Oxford Handbook of Negation*. Viviane Déprez and Teresa Espinal (eds.), 75-88. Oxford University Press.
- Munat, Judith. (2007). *Lexical creativity, texts and contexts*. John Benjamins.
- Nahajec, Lisa. (2021). *Negation, expectation and ideology in written texts: a textual and communicative perspective*. John Benjamins Publishing Company.
- Van Dijk, Teun A. (2023). (Anti)racist discourse. In *The Routledge handbook of discourse analysis*. Michael Handford and James Paul Gee (eds.) Routledge, 244-260.

Emotionality revisited

A new look at the frequency and dispersion of emotional lexis in US Hollywood film screenplays

Christian Hoffmann (Universität Augsburg, Germany)

Keywords: telecinematic stylistics, emotionality, AI-based text segmentation, key domain method, screenplays

Linguistic expressions of emotionality have been described as “the most, essential feature of telecinematic language at large, used for example to create drama and conflict” (Werner 2021: 45; Bednarek 2012, 2018). Studies have shown that emotional lexis, e.g. expressive interjections, swear words, emotional verbs, tends to be elevated in telecinematic discourse when compared to non-fictional monitor corpora of spoken (American) English (Berber Sardinha & Veirano Pinto 2017, 2019; Veirano Pinto 2018). To study the emotionality in films and television, most corpus stylistic studies have either drawn on comparatively small corpora (one feature film or television series or a limited set of episodes) or large-scale text corpora (thousands of feature films or television series based on subtitles). While the results of the first group of studies have raised concerns about their representativeness for film discourse, the second group has mainly pointed out quantitative trends in the use of emotional lexis but often failed to engage in sufficient qualitative downsampling (Werner 2021: 43).

This study aims to close this research gap, providing a closer look at the character, film or genre-based reasons for the increase of emotional features in Hollywood screenplays: Specifically designed python scripts allow for the automatic segmentation of functional text spans, based on consistent levels of textual indentation, e.g. screen directions, character tags, character contributions, etc. The procedure has been applied to create the Augsburg Screenplay Corpus (ASC, 852,659 tokens), containing 100 different US American Hollywood screenplays, equally stratified across four main film genres (Action/Thriller, Comedy, Drama, Horror) and five consecutive time periods (1970s, 1980s, 1990s, 2000s, 2010s). The ASC also holds additional annotations regarding film genres, film titles, character gender, character names and character types to systematically compare the use of emotional lexis on different levels of analysis. The study combines classic concordancing (frequency and dispersion) with the (semantic) key domain method (Archer, McEnery, Rayson & Hardie 2003). It first elicits emotional lexis in the screenplays and then explores its dispersion across different analytical strands, e.g. time, gender, film genre.

The results of this study show how, where and (potentially) why emotional expressions are employed in US Hollywood screenplays. It is shown that (a) the frequency of emotional lexis varies significantly across the four main film genres and between female and male characters, (b) the types of emotional expressions used across film genres are notably different, (c) the general level of emotionality in US Hollywood

screenplays is rather stable across time and (d) there is a considerable intrageneric dispersion of emotional lexis between films and film characters. More specifically, the overuse of expressive or emotional lexis is often linked to a very limited set of film characters, which suggests that screenwriters employ these “surge features” (Taavitsainen 1999) strategically as implicit characterisation cues (Culpeper 2001) for specific characters of their movies whilst moderating their use of for others.

References

- Bednarek, M (2012). “Get us the hell out of here”: Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics* 17(1): 35-63.
- Bednarek, Monika. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge: Cambridge University Press.
- Berber Sardinha, Tony B. and Marcia Veirano Pinto. (2017). American television and off-screen registers: A corpus-based comparison. *Corpora* 12(1): 85-114.
- Berber Sardinha, Tony B. and Marcia Veirano Pinto. (2019). Dimensions of variation across American television registers. *International Journal of Corpus Linguistics* 24(1): 3-32.
- Culpeper, Jonathan. (2001). *Language and characterisation: People in plays and other texts*. London: Routledge.
- Rayson, Paul and Nicholas Smith. (2006). The key domain method for the study of language varieties. *The Third Inter-Varietal Applied Corpus Studies (IVACS) group International Conference on Language at the Interface*.
- Taavitsainen, Irma. (1999). Personality and styles of affect in The Canterbury Tales. In *Chaucer in perspective: Middle English essays in honour of Norman Blake*. Geoffrey Lester (ed.) London: AC Black, 218-234.
- Veirano Pinto, Marcia. (2018). Variation in movies and television programs: The impact of corpus sampling. In *the language of pop culture*. Valentin Werner (ed.) London: Routledge, 139-161.
- Werner, Valentin. (2021). A diachronic perspective on telecinematic language. *International Journal of Corpus Linguistics* 26(1): 38-70.

Stop being single

A corpus-assisted discourse analysis of scholarly educational videos' comments and danmu on romantic relationships

Jing Huang (South China Agricultural University, China)

Tanjun Liu (Xi'an Jiaotong-Liverpool University, China)

Keywords: comments and danmu, romantic relationships, youth-oriented video-sharing platform, LancsBox

The rocketing divorce rate and the plummeting marriage rate in mainland China in the past few years evoked public concern about the hardships of not only maintaining a marriage but also getting into a marriage or relationship. Tuodan (literally, to get out of being single) has become “a problem”. Some academics keenly participated in such public discussions, created contents on popular social media and shared ideas or suggestions. This study focuses on a 10-episode educational collection on the topics of romantic relationships and “stop being single” uploaded on one of the most popular youth-oriented video-sharing platforms in China, Bilibili, by a wellknown university sociologist, Shen Yifei. The videos altogether led to more than 40000 comments and replies and 81000 danmu (on-screen comments like bullet curtain). Two corpora of these comments and danmu were separately built. Manual cleaning was also conducted. Drawing upon a corpus-assisted discourse approach, this study uses #LancsBox v.5.x. (Brezina, Weill-Tessier & McEnery 2020) to analyse these comments and danmu. We focused on key terms such as romantic relationships (恋爱 *lian'ai*, 谈恋爱 *tan lian'ai*), stop being single (脱单 *tuodan*) and explored their usage in collocations and contexts, examining how these concepts are constructed. The initial analysis shows that romantic relationship is commonly described in terms of patterns (模式) and disagreement has occurred on it. Some pursued after idealised or established relationship/love patterns, which is dismissed by some others, who advocate fostering realistic relationships and engaging in interactions with specific individuals. This debate corresponds to Shen's suggestion, regarding not imitating the previous generation in following up set relationship patterns. Another frequently occurring collocate of romantic relationship is self (自己 *ziji*). Many viewers emphasised their own needs and preferences over those of the dates, which indicates a self-priority approach to considering entering into a relationship. These strike a stark contrast with the Chinese previous generation, where a functionalistic approach and a couple-as-a-team orientation predominantly influenced people's decisions to enter a relationship or marriage. These findings signal an individualistic trend and resonate with the surge of egoism (Yan 2003) in relationship discourse in current China. This research contributes to understanding the transforming beliefs about love, relationships and dating in China by focusing on the comments and danmu on a youth-oriented video-sharing social platform. It helps clarify the practices involved in co-constructing romantic relationship discourses,

particularly in terms of the impact of academic influencers upon audiences as well as the agency of media participants.

References

Brezina, Vaclav, Platt Weill-Tessier and Antony McEnery. (2020). #LancsBox v. 5.x. [software]. <http://corpora.lancs.ac.uk/lancsbox>

Yan, Yunxiang. (2003). *Private life under socialism*. Stanford: Stanford University Press.

Emotivity in popular science talks?

An analysis of evaluative language used by speakers to express attitude and emotions in TED talks: Introducing the EmotionalizTED corpus

Mirela Imamovic (Universität Hildesheim, Germany)

Keywords: attitude, emotion, evaluation, popular science, TED talks

Previous studies in the identification of evaluative expressions have used Appraisal theory in the contrastive language analysis of scientific research articles (Yuchen et al. 2021), spoken language in casual conversations of various genres (Eggins & Slade 1997), translation analysis (Munday 2012a), audiovisual translation of subtitles (Munday 2012b). I apply Appraisal theory (Martin & White 2005) to the contrastive analysis of evaluative linguistic expressions in spoken popular science (PopSci) in English and German on a dataset of TED talks. The study presented here is part of a broader interdisciplinary project (a corpus-based discourse analysis of attitudinal evaluations and audience engagement) investigating *a) the use of speakers' evaluative language (attitude and emotions) in popular science speeches and the way they use audience-engaging markers (pronouns, imperatives, questions)* and *b) the translational aspect of emotions and attitude from English to German*. No attempts have yet been made to the role of these communicative functions in popular science. The study seeks to answer how the discourse topics (e.g. technology, science) are evaluated, i.e. what the most frequently expressed attitude is and if there are any domain-specific effects.

In this talk, I introduce the EmotionalizTED specialised corpus which consists of 380 texts (190 English (North American) TED talk transcripts and 190 German subtitles/translations) (see Appendix A). The data has been collected from the TED.com website. The corpus contains 95 men's and 95 women's speeches on various topics within 11 disciplines. Those disciplines are art, business, education, entertainment, history, medicine, natural science, philosophy, politics and law, psychology and technology. Selected English transcripts contain speeches of different speakers and German subtitles are all translated by different native German translators.

I report on keyword analysis findings on the use of speaker self-mentions (an aspect of audience engagement, see Hyland 2005) in all English transcripts. The reference corpus used was the COCA (spoken) corpus. The analysis shows that the pronoun "we" is used in distinctive patterns followed by the emotion verbs (*we all + verb*) and followed by evaluative adjectives (*we are + adjective*), e.g. "*we all feel*" and "*we are powerless*". The results suggest that the speakers' tendency to use pronouns (referring directly to themselves or people in general) as a way to engage the audience is also a mechanism for expressing their attitudes. The analysis of German translations reveals that "*we all*" and "*we are*" are directly translated as "*wir alle*" and "*wir sind*", respectively. The use of exact patterns in German translations remains to be analysed further. Future work will also include an analysis of more fine-grained evaluative linguistic patterns in

English and compare the expressions of emotion and attitude in the corpus of TED talks (spoken PopSci) with a corpus of PopSci written articles. The evaluative expressions in English will then be compared to the ones in German translations to find out whether they are translated directly, or if they are altered or omitted.

References

- Egins, Suzanne and Diana Slade. (1997). *Analysing casual conversation*. London: Cassell.
- Hyland, Ken. (2005). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies* 7(2): 173-191.
- Martin, James and Peter R. R. White. (2005). *The language of evaluation*. New York: Palgrave Macmillan.
- Munday, Jeremy. (2012a). New directions in discourse analysis for translation: A study of decision-making in crowdsourced subtitles of Obama's 2012 State of the Union speech. *Language and Intercultural Communication* 12(4): 321-334.
- Munday, Jeremy. (2012b). *Evaluation in translation: Critical points of translator decision-making*. London: Routledge.
- TED.com. Accessed on 28 July 2023. <https://www.ted.com/>
- Yuchen, Xu, Yan Xuan, Su Rui and Kou Ying. (2021). *A corpus-based contrastive study of the appraisal systems in English and Chinese scientific research articles* (1st ed.). London: Routledge.

Appendix

	EmotionalizTED Corpus			
DISCIPLINE	ENGLISH ORIGINALS		GERMAN TRANSLATIONS	
	Transcripts	Words	Subtitles	Words
Art	16	35,973	16	31,753
Business	16	32,639	16	29,209
Education	15	37,050	15	32,010
Entertainment	15	30,269	15	27,408
History	15	29,242	15	26,334
Medicine	17	41,636	17	37,936
Natural Science	22	55,468	22	49,280
Philosophy	15	41,841	15	38,730
Politics and Law	17	42,536	17	38,482
Psychology	17	39,567	17	35,127
Technology	25	43,209	25	37,960
11	190	429,430	190	384,229
Total texts: 360				
Total words: 813,659				

Negotiating metaphorical frames:

A corpus-assisted study of sinophobic discourses in editorials and reader comments during the COVID-19 pandemic

Ilaria Iori (Università degli Studi di Modena e Reggio Emilia, Italy)

Keywords: metaphor, sinophobia, news discourse, reader comments, discriminatory discourses

Studies on representations of migrants have underscored the crucial role of metaphors in discriminatory discourses (Musolff 2015; Taylor 2021). Recent investigations on COVID-19-related Sinophobia observed negative evaluative frames in different media platforms, including social media (Lee 2021) and news websites (Cho 2021; Kim, Cooks, & Kim 2022). Particularly in news websites, the play of voices in the comment sections can be particularly insightful, as readers can extend or twist the communicative purpose of the metaphorical frames first proposed in the editorials (A'Beckett 2012). Despite these aspects, little attention has been paid in the literature to the extent to which metaphors in news articles and editorials are reproduced in their respective comment sections. Hence, this study explores the textual voices emerging from the reader comment sections taking a dialogic perspective. Specifically, it investigates the communicative functions of metaphors used to frame China and Chinese social actors in editorials and reader comments, concentrating on how metaphors first proposed in the editorials are reproduced in the comment sections. The case study focuses on one of the most-read newspapers in Italy (*La Repubblica*) and Australia (*Daily Telegraph*), as both countries experienced significant cases of anti-Asian hate (see Kamp et al. 2023; Miyake 2021). The analysis addresses the following research questions: (i) What are the most recurrent metaphors used to represent China and Chinese social actors in the editorials? (ii) To what extent are these metaphors reproduced in the comment sections? (iii) What functions do they perform in editorials and reader comments? Methodologically, the study combines discourse approaches and corpus linguistic methods (Partington, Duguid & Taylor 2013). An ad hoc corpus was compiled collecting all editorials on China published in the first six months of 2020 with their respective comment sections. Drawing on Charteris-Black (2004), candidate metaphors were manually recognised in a smaller sample of texts and their concordance lines were closely analysed using #Lancsbox 6 (Brezina, WeillTessier, & McEnery 2020). Metaphors were identified referring to the Metaphor Identification Procedure (Pragglejaz Group 2007). Then their functions were examined referring to the Conceptual Metaphor Theory (Lakoff & Johnson 1980). Results show that metaphors in the editorials frequently carried negative connotations in both the Australian and Italian corpus (e.g., 'monster' or 'impero senz'anima'/empire without a soul). However, in the Italian editorials, more positive metaphors were also recognised (e.g., 'amico'/friend). Australian reader comments tended to reproduce and expand upon the metaphorical scenarios introduced in the editorials using various linguistic expressions. In the Italian comments, metaphors were reproduced to extend the meaning of the metaphorical

frame or strongly criticise it. Despite some cross-cultural variations, several metaphors discursively constructed Orientalist representations of China and its social actors in both the Australian and Italian corpus, portraying them as the exotic other (Kawai 2005). This framing contributed to creating an 'us versus them' narrative (Van Dijk 1998), potentially stigmatising the social actors involved.

References

- A'Beckett, Ludmilla. (2012). The play of voices in metaphor discourse: A case study of "NATIONS ARE BROTHERS". *Metaphor and Symbol* 27(2): 171-194. 10.1080/10926488.2012.665797
- Brezina, Vaclav, Will-Tessier, Pierre and Tony McEney. (2020). #Lancsbox v.5.x.' [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>
- Charteris-Black, Jonathan. (2004). *Corpus approaches to critical metaphor analysis*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Cho, Eunil David. (2021). From the yellow peril to the model minority and back again: Unraveling the orientalist representations of Asian Americans in the age of covid-19'. *Journal of Pastoral Theology* 31(2-3): 175-192. 10.1080/10649867.2021.1929711
- Kamp, Alanna, Sharples, Rachel, Vergani, Matteo and Nida Denson. (2023). Asian Australian's experiences and reporting of racism during the COVID-19 pandemic. *Journal of Intercultural Studies*: 1-21. 10.1080/07256868.2023.2290676
- Kawai, Yuko. (2005). Stereotyping Asian Americans: The dialectic of the model minority and the yellow peril. *Howard Journal of Communications* 16(2): 109-130. 10.1080/10646170590948974
- Kim, Bumsoo, Cooks, Eric and Kim Seong-Kyu. (2022). Exploring incivility and moral foundations toward Asians in English-speaking tweets in hate crime-reporting cities during the COVID-19 pandemic. *Internet Research* 32(1): 362-378. 10.1108/INTR-11-2020-0678
- Lakoff, George and Mark Johnson. (1980). *Metaphors we live by*. Chicago: University of Chicago Press.
- Lee, Carmen. (2021). #HatelsAVirus: Talking about COVID-19 "Hate". In *Viral discourse*. Rodney H. Jones (ed.) Cambridge Elements. Elements in Applied Linguistics. Cambridge: Cambridge University Press.
- Miyake, Toshio. (2021). "Cin Ciun Cian" (Ching Chong): Yellowness and neo-orientalism in Italy at the time of COVID-19. *Philosophy & Social Criticism* 47(4): 486-511. 10.1177/01914537211011719
- Musolff andreas. (2015). Dehumanizing metaphors in UK immigrant debates in press and online media. *Journal of Language Aggression and Conflict* 3(1): 41-56. 10.1075/jlac.3.1.02mus.
- Partington, Alan, Duguid, Alison and Charlotte Taylor. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)* 55. Amsterdam: John Benjamins Publishing Company. 10.1075/scl.55
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1): 1-39. 10.1080/10926480709336752
- Taylor, Charlotte. (2021). Metaphors of migration over time. *Discourse & Society* 32(4): 463-481. 10.1177/0957926521992156
- van Dijk, Teun A. (1998). *Ideology: A multidisciplinary approach*. London; Thousand Oaks; New Delhi: Sage Publications.

“That’s diabetes on a plate”

Analysing online news reader comments about the representation of diabetes on The Great British Bake Off

Ursula Kania (University of Liverpool, United Kingdom)

Antonia Phillips-Beeching (University of Liverpool, United Kingdom)

Keywords: corpus-assisted critical discourse analysis, online news media, diabetes, health communication, participatory online spaces

During the season 10 semi-final of the UK TV show *The Great British Bake Off* (henceforth: GBBO), judge Paul Hollywood commented that a contestant’s bake looked like ‘diabetes on a plate’, triggering complaints by viewers as well as a broader debate on media representations of diabetes.

The fact that media representations of diabetes matter (and that misrepresentations have negative effects such as increasing stigma for those affected) has long been acknowledged (Hunt et al. 2022, for a recent summary/study). However, even though GBBO has attracted some academic attention (Bradley 2016), its role in both (allegedly) mirroring as well as shaping public opinion not only on conditions such as diabetes but also regarding various other aspects of ‘health’ and ‘diet culture’ has not been scrutinized systematically yet.

The current study is a first step towards closing this gap, using the debate around Paul Hollywood’s comment on diabetes as a case study. As part of the UK media coverage of the incident, the tabloid *Daily Mail* published an article on 23th October, 2019, entitled ‘Paul Hollywood apologises after GBBO fans branded him ‘vile’ for likening a cake to ‘diabetes on a plate’’. Like many other online news outlets, the *Daily Mail* invites readers to submit comments – since these comment sections “provide news audiences with an immediate, interactive and far-reaching possibility to participate in the interpretation of current events” (Ziegele 2019: 1), they offer valuable insights into readers’ perceptions and attitudes related to the news story.

All 863 user comments posted under the article were extracted from the *Daily Mail* website (using the free version of *Data Scraper*) and the resulting corpus (approx. 17,000 words) was subjected to quantitative as well as qualitative analyses (e.g., keywords, word frequencies, sentiment analysis; supported by *AntConc* and *NVivo*), using corpus-assisted critical discourse analysis (Baker & McEnery 2015).

Findings show that comments fall into three different (sometimes overlapping) themes: 1. health/lifestyle, 2. media/media personalities, 3. ‘woke culture’. While the presentation of findings will focus on the first two themes and associated discourses (such as ‘Type 2 diabetes as self-inflicted due to poor individual lifestyle choices’ and ‘Media personalities should be positive role models’), mentions of ‘woke culture’ and closely associated terms (such as ‘snowflake(s)’) will also be discussed, due to their prevalence. Now often removed from its original meaning (in the struggle for civil rights), ‘woke’ can

be mobilised as an insult, “signify[ing] a supposed progressive over-reaction” (Cammaerts 2022: 735), to shut down any discussion about discrimination or problematic representations by stating that ‘it was just a joke’. Comments within this theme are discussed in the context of specific affordances provided by this participatory online space (such as upvoting and downvoting other contributions, replying to other users, etc.).

Overall, this study contributes to our understanding of (online news readers’ discussion of) media representations of diabetes (and other health issues) and provides insights into the dynamics of comments sections provided by online news media.

References

- Baker, Paul and Tony McEnery. (2015). *Corpora and discourse: Integrating discourse and corpora*. London: Palgrave.
- Bradley, Peri. (2016). More cake please — We’re British! Locating British identity in contemporary TV food texts, the great British bake off and come dine with me. In *Food, Media and Contemporary Culture*. Peri Bradley (ed.) London: Palgrave Macmillan. [10.1057/9781137463234_2](https://doi.org/10.1057/9781137463234_2)
- Cammaerts, Bart. (2022) The abnormalisation of social justice: The “anti-woke culture war” discourse in the UK. *Discourse & Study* 33(6): 719-846. [10.1177/09579265221095407](https://doi.org/10.1177/09579265221095407)
- Hunt, Daniel, Lamb, Krista, Elliott, James, Hemmingsen, Bianca, Slama, Slim, Scibilia, Renza, Whitney Daniels, Kristen and Bente Mikkelsen. (2022). A WHO key informant language survey of people with lived experiences of diabetes: Media misconceptions, values-based messaging, stigma, framings and communications considerations. *Diabetes Research and Clinical Practice* 193. [10.1016/j.diabres.2022.110109](https://doi.org/10.1016/j.diabres.2022.110109)
- Ziegele, Marc. (2019). Reader commenting. In *The international encyclopedia of journalism studies*, 1-8. [10.1002/9781118841570.iejs0059](https://doi.org/10.1002/9781118841570.iejs0059)

Business knights and superwomen

Gendered representation of leaders in Estonian media.

Elisabeth Kaukonen (University of Tartu, Estonia)

Raili Marling (University of Tartu, Estonia)

Keywords: Estonian, media discourse, leadership, gender studies

Occupations, especially leadership, continue to be viewed through gender ideologies. Numerous studies have shown, for instance, that women are judged to be unfit for leadership positions due to gender stereotypes (Heilman 2001; Koburtay et al. 2018), as authority still continues to be associated with masculinity (Schein 1973; Castaño et al. 2019). As a result, it is harder for women to get promoted and they remain under-represented in top leadership positions (Cook & Glass 2014). This situation has not dramatically improved, despite decades of research and policy initiatives. These stereotypes persist in public and media discourse. While this topic has been studied extensively, there is only limited research on the post-socialist world (for Estonia, Pilvre 2011). Yet countries like Estonia are interesting case studies, as post-Soviet suspicion of gender equality initiatives coexists with EU legislation and egalitarian practices of Nordic companies. This complex discursive blend merits closer attention.

This presentation grows out of a collaborative project of Praxis Think Tank and University of Tartu which studies gender and organizational culture in private and public sector enterprises. The presentation specifically focuses on one section of this study, gendered discourses in the media coverage of male and female leaders in Estonia. In the study, web subcorpus of Estonian National Corpus 2021 (Koppel & Kallas 2022), as well as interviews and profile stories from two Estonian media publications were studied with the help of corpus-assisted discourse studies and critical discourse analysis (Baker 2006; Partington et al. 2013; Stefanowitsch 2020). The aforementioned corpus includes different web texts, such as news articles, homepages, as well as forums, blogs, commentaries and many more. Interviews and profile stories were collected from a media database Station. Our study aimed to find answers to following research questions:

- Through which discourses are Estonian female and male leaders portrayed?
- What types of femininity and masculinity are highlighted in the case of leaders?
- What leadership qualities do the leaders mention and how are they gendered?
- How is the leader's gender conveyed in media texts?

The results demonstrate that both female and male leaders are portrayed through gender stereotypes, as either ultra-feminine or ultra-masculine, respectively. The results confirm previous research, according to which, for example, female leaders are often portrayed as superwomen, women face more prejudice in leadership positions and it is assumed that the lack of female leaders is often women's own fault. (see Sheerin &

Garavan 2021) However, the material also reflected a shift towards valuing leaders who exhibit traits conventionally associated with femininity (Koburtay et al. 2019).

References

- Baker, Paul. (2006). *Using corpora in discourse analysis*. London/New York: Continuum.
- Castaño, Ana M., Fontanil, Yolanda and Antonio L. García-Izquierdo. (2019). Why can't I become a manager? - A systematic review of gender stereotypes and organizational discrimination. *International Journal of Environmental Research and Public Health* 16(10): 1-29. [10.3390/ijerph16101813](https://doi.org/10.3390/ijerph16101813)
- Cook, Alison and Christy Glass. (2014). Women and top leadership positions: towards an institutional analysis. *Gender, Work and Organization* 21(1): 91-103. [10.1111/gwao.12018](https://doi.org/10.1111/gwao.12018)
- Heilman, Madeline E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues* 57(4): 657-674. [10.1111/0022-4537.00234](https://doi.org/10.1111/0022-4537.00234)
- Koburtay, Tamer, Jawad, Syed and Radi Haloub. (2019). Congruity between the female gender role and the leader role: A literature review. *European Business Review* 31(6): 831-848. [10.1108/EBR-05-2018-0095](https://doi.org/10.1108/EBR-05-2018-0095)
- Koppel, Kristina and Jelena Kallas. (2022). Eesti keele ühendkorpuste sari 2013–2021: mahukaim eestikeelsete digitekstide kogu [Estonian National Corpus 2013–2021: the largest collection of Estonian language data]. *Eesti Rakenduslingvistika Ühingu Aastaraamat* 18: 207-228. [10.5128/ERYa18.12](https://doi.org/10.5128/ERYa18.12)
- Partington, Alan, Duguid, Alison and Charlotte Taylor. (2013). *Patterns and meanings in discourse. Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Pilvre, Barbi. (2011). Naiste meediarepresentatsioon Eesti ajakirjanduskultuuri ja ühiskonna kontekstis. [Media representation of women in the context of Estonian journalist culture and society]. *Dissertationes de Mediis et Communicationibus Universitatis Tartuensis* 12. Doctoral thesis. <https://dspace.ut.ee/server/api/core/bitstreams/f945a9ab-e8484fbb-a224-d884abdd269d/content>
- Schein, Virginia E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology* 57(2): 95-100. [10.1037/h0037128](https://doi.org/10.1037/h0037128)
- Sheerin, Corina and Thomas Garavan. (2021). Female leaders as 'superwomen': Post-global financial crisis media framing of women and leadership in investment banking in UK print media 2014-2016. *Critical Perspectives on Accounting* 68: 1045-2354 [10.1016/j.cpa.2021.102307](https://doi.org/10.1016/j.cpa.2021.102307)
- Stefanowitsch, Anatol. (2020). *Corpus linguistics. A guide to methodology. Textbooks in Language Sciences*. Berlin: Language Science Press.

Using corpus linguistics to safeguard children against online harms

Charlotte-Rose Kennedy (Birmingham City University, United Kingdom)

Keywords: safeguarding, EdTech, Knowledge Transfer Partnership, applied corpus linguistics

Funded by Innovate UK, the Knowledge Transfer Partnership (KTP) is a UK-wide programme that connects forward-thinking businesses with the UK's world-class knowledge bases to deliver business-led innovation projects. This paper focuses on findings from the KTP between Birmingham City University and Renato Software Limited, a company specialising in a cloud-based classroom management and safeguarding software called 'Senso.cloud'. Senso.cloud currently monitors students' keystrokes to identify keyword violations that signal potential safeguarding issues (e.g., 'bomb', 'kill myself') so that schools' Designated Safeguarding Leads can intervene and protect children. The KTP seeks to enhance Senso.cloud's keyword violation detection, utilising CADS methods (Partington et al. 2013) to improve the context sensitivity and overall accuracy of the monitoring system.

To achieve this, a 4,326,947-word corpus (comprising 6,448 documents) of testimonies concerning safeguarding issues was collected from a range of online sources, including mental health blogs; charity websites; Reddit; X (Twitter); and YouTube. The overall dataset encompasses five sub-corpora that relate to specific safeguarding themes. These include suicide; mental health and eating disorders; bullying and discrimination; drugs and substance abuse; and sexual exploitation. This paper focuses on the 'suicide' sub-corpus, comprising 1,094,914 words (25.3% of the total dataset).

Using Sketch Engine's (Kilgarriff et al. 2014) keyword tool, the 'suicide' sub-corpus was compared to a 14,461,253-word corpus of children's Microsoft Teams messages to identify statistically significant key terms. Sketch Engine's word sketch and concordance tools were then used to uncover the contexts in which the key terms categorised as 'mental processes' and 'adverbs' were embedded. By using CADS methods to count what is difficult to count (Marchi 2024), the findings of this analysis revealed some of the more subtle, non-explicit phrases children might type on school computers that indicate they are suffering from mental health issues and suicidal ideation. These phrases will be added to Senso.cloud's keyword violation libraries, enhancing the system's detection of mental health issues and potentially saving lives.

As well as helping Renato Software Limited to become the most linguistically accurate safeguarding tool in the EdTech industry, positioning the company as a market leader, the results of the linguistic analysis conducted throughout the KTP will inform the UK Government's response to online threats. This will ultimately improve the safeguarding capabilities of institutions bound by a duty of care (i.e. schools) and help to protect children from online harms.

References

Kilgarriff, Adam, Baisa, Vit, Bušta, Jan, Jakubíček, Milos, Kovář, Vojtech, Michelfeit, Jan, Rychlý, Pavel and Vit Suchomel. (2014). The sketch engine: ten years on. *Lexicography* 1: 7-36.

Marchi, Anna. (2024). Get back! Methods for exploring discourses of nostalgia and nostalgic discourse using corpora. *Elephant & Castle* 31: 192-211.

Partington, Alan, Duguid, Alison and Charlotte Taylor. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins.

Turn and turn again

On efficiency and formulaity in humorous fictional TCD

Monika Kirner-Ludwig (Universität Innsbruck, Austria)

Keywords: formulaic humor, humorous formulae, telecinematic dialogue, fictional-conversation analysis (FCA), popcultural linguistics

In alignment with the conference focus, this study offers insights into the text type of telecinematic dialogues (e.g. Bednarek 2018, 2017; Piazza et al. 2011: 1) and how they are structured and patterned when they specifically feature humorous formulae (cf. Kirner-Ludwig 2023). Such humorous formulae may range from humorous interjections (e.g. <How about them X>), proverb-like formulae (e.g. <it takes one to know one>) and situation-bound utterances (e.g. <not if I see you first>) to various recitable joke types (e.g. <your mama>; <X called they want their Y back>) and isolated formulaic punchlines (e.g. <that's what she said>; cf. Kirner-Ludwig 2018, 2023). Samples are taken from a self-compiled Sketch Engine corpus, comprising 19,422,645 words from 34 US American TV series.

While much linguistic discussion has been dedicated to “the limitations and gains of using [instances of scripted] fiction [...] as specimens for the study of language” (AlvarezPereyre 2011: 47; also cf. Attardo 2001) and to what extent “artistic-commercial films [and other scripted TCD might] prove ‘good enough’ substitutes” for real spontaneous interactions (Alvarez-Pereyre 2011: 49), this study is going to zoom in on fictional TCD’s unique features, i.e. on the fact that, when it comes to formulaic humor, the architecture and construction of fictional TCD dialogues is inherently and recurrently pattern-like and, in that, will significantly outrun spontaneous unscripted speech in terms of expectability of formulaic turns. This will be shown by applying a CA-transcription template that – on the basis of earlier case studies – itemizes expected turn types (_exp) and optional (sets of) turn types (_opt) in dialogues featuring HFs.

References

- Bednarek, Monika. (2017). The role of dialogue in fiction. In *Pragmatics of Fiction (Handbooks of Pragmatics 12)*. Miriam A. Locher and Andreas H. Jucker (eds.) Berlin; Boston: De Gruyter Mouton, 129–158.
- Bednarek, Monika. (2018). *Language and television series: A linguistic approach to TV dialogue*. Cambridge: Cambridge University Press.
- Kirner-Ludwig, Monika. (forthcoming). *Humorous formulae and formulaic humor*. Amsterdam: John Benjamins.
- Piazza, Roberta, Bednarek, Monika and Fabio Rossi. (2011). *Telecinematic discourse: Approaches to the language of films and television series*. Amsterdam: John Benjamins.

Metaphors of slavery

A metaphor analysis of pro- and anti-slavery pamphlets

Laurin Klocker (Universität Innsbruck, Austria)

Keywords: metaphors, discourse analysis, political discourse, slavery, abolitionism

Emerging in the 1780s, the British abolitionist movement sought to dismantle the institutions of the African slave trade and ameliorate the conditions of the enslaved in the colonies. The debate dragged on for decades, owing to the fierce opposition mounted by slavery interest groups. Among other outlets, pamphlets served as an important tool to spread information and sway public opinion (Dumas 2016: 51-56). The present study focuses on the metaphors used to frame the enslaved and members of the slavery interest within such pamphlets, building on the theoretical foundation of conceptual metaphor analysis laid by Lakoff and Johnson (1980). These metaphors can highlight the conceptual frameworks underlying the respective political stances and reveal the strategies used to justify or vilify the slave trade (Charteris-Black 2011: 36; Lakoff & Johnson 1980: 3-5).

For the analysis, a corpus compiled of three pamphlets from each of the opposing political camps was created and analyzed with the AntConc corpus processor. As the core issues of the debate shifted over the years (Dumas 2016: 12), only texts published during the early efforts of the abolitionist movement were considered to ensure comparability. Keywords from the respective target domains served as the basis for the identification of relevant metaphors, with the data thus gathered being supplemented with a close reading of the pamphlets.

The metaphors reveal both similarities and differences in the framing of the enslaved and the slavery interest. While OBJECT and ANIMAL metaphors were frequently employed on both sides, they may either serve to dehumanize or to create sympathy for the enslaved. Proposals to ameliorate the conditions on plantations resulted in a shift from the OBJECT to the ANIMAL frame: The enslaved were to be bred and treated like LIFE STOCK, as opposed to their being viewed a mere RESOURCE to be purchased and expended. Members of the interest, on the other hand, were framed as CRIMINALS by abolitionists, while their opposition sought to reverse the role of victim and oppressor. Most strikingly, although the abolitionist pamphlets show a greater tendency towards humanizations, many of their metaphors still reflect the deep-seated notions of racial hierarchy and differentiation of the time.

References

Charteris-Black, Jonathan. (2011). *Politicians and rhetoric: The persuasive power of metaphor*. 2nd ed. Basingstoke: Palgrave Macmillan.

Dumas, Paula E. (2016). *Proslavery Britain: Fighting for slavery in an era of abolition*. New York: Palgrave Macmillan.

Lakoff, George and Mark Johnson. (1980). *Metaphors we live by*. Chicago: Chicago University Press.

A classification scheme of discourse studies on social media

The benefits and pitfalls of adopting a CADS approach

Susanne Kopf (Wirtschaftsuniversität Wien, Austria)

Keywords: social media, social media analysis classification scheme, corpus-assisted discourse studies

This paper deals with how social media and social media data can be approached via corpus-assisted discourse studies (CADS). It introduces a classification scheme for the different ways of studying social media from a discourse analytical angle. For this, it proposes that, in the main, discourse analytical research can address social media from four different angles. In addition, the paper details how these four approaches to studying social media may benefit specifically from a Corpus-Assisted Discourse Studies (CADS) methodology. It also discusses potential weaknesses of taking a CADS approach and how these weaknesses can be mitigated.

Social media, here defined as “web-based and mobile services that allow individuals, communities and organizations to collaborate, connect, interact and build community by enabling them to create, co-create, modify, share and engage with content (user- or bot-generated)” (Nau, Quan-Haase & McCay-Peet 2022: 15), have yielded ample research in linguistics generally and discourse studies specifically. Principally consisting of digital language material (in addition to other forms of semiosis), social media data lend themselves to collecting and corpus building. That is, at least in comparison to analogous data which need to be digitised, etc. first, building corpora from social media data is relatively easy and there are even tools that support automated scraping or annotation of social media data (Berberich & Kleiber 2023 for an overview of various tools). Therefore, it is unsurprising that past research in corpus linguistics has dealt with social media and social media data, ranging from, for example, corpus pragmatic studies on business to customer interaction on Twitter (Lutzky 2021) to corpus-assisted discourse analyses of governmental social media communication (Hansson & Page 2023). Indeed, even the issue of social media’s multimodal in addition to linguistic meaning-making has received research attention in corpus-assisted discourse studies, see e.g. Corpus-Assisted Multimodal Discourse Analysis (CAMDA) (see more on this in, e.g. Caple 2018).

This paper complements existing research by introducing a classification scheme of how social media can be studied from a discourse analytical perspective. Specifically, it contends that past linguistic and discourse-analytical research falls into four broader categories: 1) Social media as data repositories, 2) Discourse(s) about social media, 3) Discourse by social media providers and 4) Discourse analytical contextualisation and theorisation of social media. After detailing these categories, the presentation explores how the categories established can be studied through a CADS lens. Here, the paper addresses the benefits of taking a corpus-assisted discourse analytical approach to social media, e.g. mitigating the danger of cherry-picking data. In addition, it discusses potential

pitfalls of adopting a CADS perspective and deals with how to avoid certain issues when designing a corpus-assisted discourse analytical study of social media.

References

Berberich, Kristin and Ingo Kleiber. (2023). *Corpus analysis: Tools for corpus linguistics*.

Caple, Helen. (2018). Analysing multimodal text. In *Corpus approaches to discourse: A critical review*. Charlotte Taylor and Anna Marchi (eds.) Routledge, 85-109.

Hansson, Sten and Ruth Page. (2023). Legitimation in government social media communication: The case of the Brexit department. *Critical Discourse Studies* 20(4): 361-378. 10.1080/17405904.2022.2058971

Lutzky, Ursula. (2021). *The discourse of customer service tweets: Planes, trains and automated text analysis*. London; New York: Bloomsbury Academic.

Nau, Charlotte, Quan-Haase, Anabel and Lori McCay-Peet. (2022). Defining social media and asking social media research questions: How well does the Swiss army knife metaphor apply? In *the SAGE handbook of social media research methods*. Anabel Quan-Haase and Luke Sloan (eds.) London: Sage, 13–26.

The evolution of woke

From social activism to political propaganda

Arina Kurkina (ITMO University, Russian Federation)

Keywords: woke, polysemy, context, corpus, dictionary

For the past few years, the word *woke* has been a focal point of the political conversation revolving around the clash of values in present-day America.

Originating in the African-American community, the adjective *woke* was initially used as a means of drawing attention to racial discrimination by left-wing activists. However, it is now being met with increasing derision and condemnation from both conservatives and liberals. The strategic use of *woke* as a tool of political propaganda not only serves to discredit opponents, but also amplifies the polarization of public discourse, effectively weaponizing language. (Sobande, Kanai & Zeng 2022; Kóczé, 2022)

That said, contemporary online English dictionaries (Collins, Merriam-Webster and Cambridge) fail to present any changes in the word meaning, describing the adjective as a primarily positive phenomenon. In order to clarify the meaning of the adjective *woke*, 400 randomly selected contexts from American media websites for the periods from 2013 to 2017 and from 2022 to 2023 were analysed. The selection criteria for media outlets, encompassing FOX News, CNN, NBC and The New York Times, were based on the ideological spectrum as outlined by The Pew Research Center¹, thereby ensuring a comprehensive overview of both conservative and liberal perspectives. The analysed contexts were restricted to the contexts with *woke* as an adjective, resulting in a corpus of 26,000 words.

The context analysis revealed two senses of *woke* characterized by contrasting evaluation.

The first sense (*negative woke*) refers to the ideology of the progressive wing of the Democratic party as a form of toxic ideology that the Democrats forcibly promote in educational institutions, government structures and in the armed forces; *woke* is represented as an imminent threat to the values of America, with such words as *indoctrination*, *infiltrate*, *tyranny* occurring in the near context.

- *He [Rep. Jim Banks] added: "The most toxic part of this tyranny is its doctrine - 'wokeness.'" "So, he's taking action: 'This Congress, I will create the first-ever AntiWoke Caucus.'" "Wokeness is especially prevalent and dangerous at universities and in primary schools because the Left recognized students as the most vulnerable and useful targets for indoctrination."* (FOX News)

¹ URL: <https://www.pewresearch.org/journalism/2020/01/24/americans-are-divided-by-party-in-these-sources-they-turn-to-for-political-news/>

The second sense (*positive woke*), on the contrary, highlights how *woke* ideology promotes critical thinking, active citizenship and initiative, all considered essential qualities for the American identity.

- *Actually, the U.S. military is sort of woke, in the sense that it is highly diverse and inclusive, encourages independent thinking and initiative on the part of junior officers and is, at the higher levels, quite intellectual.* (The New York Times)

The distribution of context types across time periods shows that contexts using *woke* from Republican perspectives are the most prevalent at the moment (275 out of 300 contexts). In contrast, the years 2013 to 2017 saw a prevalence of the positive interpretation of the word *woke* (95 out of 100 contexts). It can be assumed that initially the adjective *woke* had an exclusively positive evaluative potential and was used primarily by activists and in media devoted to pop culture. Thus, the Republicans managed to appropriate this adjective and use it as a political weapon. The original positive sense of the word has been forced out of use.

References

Sobande, Francesca, Akane Kanai and Natasha Zeng. (2022). The hypervisibility and discourses of 'wokeness' in digital culture. *Media, culture & society* 44(8): 1576-1587.

Kóczé, Angela. (2022). Dual attack on social justice: the delegitimization of "woke"/"wokism" //. *Culture Wars Papers* 31.

Actualization between exemplar-based analogy and recontextualization

Up in British English and American English during the years 1600-1999

Ljubica Leone (Università degli Studi di Milano Statale, Italy)

Keywords: *up*, actualization, recontextualization, analogy, varieties

The English word *up* is characterized by a complex linguistic behavior that derives from its multifunctional substance. According to modern grammars (Quirk et al. 1985; Biber et al. 1999) *up* can be a preposition working as the head of a prepositional phrase (e.g. *up the hill*), or it can behave as an adverb specifying place and direction (Bolinger 1971) which, on many occasions, is paradigmatically linked to *upstairs* (e.g. *it is up/upstairs*); At the same time, it may occur as an adverbial particle in phrasal verbs (e.g. *to give up*) or in phrasal-prepositional verbs (e.g. *to live up to*) (Denison 1985). The multifunctional substance and the existence of a great amount of variability in the use of *up* that is registered in modern times are the results of interacting factors that prompted the linguist renewal of this form over time.

Existing studies have demonstrated that *up* underwent syntactic reanalysis driven by exemplar-based analogy and was affected by recontextualization intended as the process resulting in “new options for the re-use of a linguistic construction” (von Mengden & Kuhle 2020: 253) which favored divergent paths of development and phenomena of layering. The actualization of the multifunctional status of *up* has been linked to processes of change leading it from the status of a free preposition or an adverb to a bounded status: syntactic reanalysis and analogy transformed *up* into a particle occurring in phrasal verbs and phrasal-prepositional verbs. These processes characterized the history of *up* since earlier periods of English (Denison 1981; Hiltunen 1983) and were particularly prominent in the Early Modern English (EModE) time (Claridge 2000). However, despite extensive research there are no studies that describe “the order in which new uses appear” (De Smet 2012: 601) in BrE. Moreover, they have not described the extent to which it is possible to generalize the observed uses of *up* from British English (BrE) to other varieties such as American English (AmE).

The present study aims to fill this gap and to discuss changes driven by exemplar-based analogy and recontextualization by taking *up* as a case study. Specifically, the aims are: (1) to describe processes of actualization characterizing *up* during the years 1600-1999 in BrE; (2) to examine divergences in AmE and BrE over time.

The present research is a corpus-based investigation undertaken on the ARCHER corpus, which is a multi-genre corpus that is available online on the CQPweb server (Hardie 2012). Concordance-based analysis of instances has been matched to quantitative evaluation.

The analysis reveals that: 1. the process of actualization of *up* as an adverbial particle may be seen as of a four-stage path leading prepositions and adverbs to the

bounded status of a particle; 2. new instances emerged via exemplar-based analogy, whereas well-established verbs underwent meaning recontextualization; 3. there are limited divergences between AmE and BrE over the years 1750-1999 in terms of productivity, but there are signs of phenomena of cross-varietal easement favouring the reciprocal influence of these two varieties.

References

- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan and Edward Finegan. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited.
- Bolinger, Dwight. (1971). *The phrasal verb in English*. Cambridge: Harvard University Press.
- Claridge, Claudia. (2000). *Multi-word verbs in early modern English. A corpus-based study*. Amsterdam and Atlanta, GA: Rodopi.
- De Smet, Hendrik. (2012). The course of actualization. *Language* 88(3): 601-633.
- Denison, David. (1981). *Aspects of the history of English group-verbs, with particular attention to the syntax of the ORMULUM*. Oxford: University of Oxford Ph.D. Dissertation. <http://www.escholar.manchester.ac.uk/uk-ac-man-scw:74782>.
- Denison, David. (1985). The origin of completive *up* in English. *Neuphilologische Mitteilungen* 86:37-61.
- Hardie andrew. (2012). CQPweb—combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 173: 380-409.
- Hiltunen, Risto. (1983). *The decline of the prefixes and the beginnings of the English phrasal verb: The evidence from some Old and Early Middle English texts*. Turku: Turun Yliopisto.
- Quirk, Randolph, Greenbaum, Sydney, Leech, Geoffrey and Jan Svartvik. (1985). *A comprehensive grammar of the English language*. Harlow: Longman.
- von Mengden, Ferdinand and Anneliese Kuhle. (2020). Recontextualization and language change. *Folia Linguistica Historica* 41(1): 253-281.

But what about jealousy?

Emotion talk in news discourse about polyamory

Alon Lischinsky (Oxford Brookes University, United Kingdom)

Keywords: news discourse, ethical non-monogamy, emotion, intimate citizenship, stigma

Contemporary Western discourse about intimate relationships is dominated by compulsory monogamy (Barker & Langdridge 2010; Ferrer 2018; Schippers 2016). As a result, people who openly form non-monogamous attachments face significant stigma and rejection and such relationships are widely believed to be immature, emotionally stunted and riddled by conflict (Moors et al. 2021; Séguin 2019).

Scholars and activists have often pointed out the importance of the media in reproducing this stigma (Antalffy 2011; Barker 2005; Klesse 2016; Séguin 2019). Ritchie & Barker (2006: 596) argue that the “emotions of polyamorous people are not easily described through the dominant language of monogamy”, for example the narratives tropes — such as the love triangle — that implicitly treat experiences of nonmonogamous desire as threatening and inherently unstable (Saxey 2010: 28). But despite the increased visibility of nonmonogamous relationships in entertainment and news media (Adams & Rambukkana 2018; Barker & Langdridge 2010: 749; Hurson 2016: 14; Moors 2017: 677), there are yet no empirical examinations of how journalistic coverage portrays, explains and evaluates the emotional dimension of polyamory.

In this paper, I seek to describe what emotions are discussed in news discussions of polyamory and how they are characterised. Using the Factiva database (Dow Jones & Company 2024), I retrieved all articles from newspaper sources containing the word ‘polyamory’ or related terms, a total of 9340 documents covering the period 1994–2023. The articles were downloaded from the database and converted using the Python BeautifulSoup library (Richardson 2023) into structured text files containing the full text and headline as well as metadata including date, country, publication, medium (online or print) and byline.

Using the Python version of the UCREL Semantic Analysis System (Moore & Rayson 2022), I identified and tagged all terms and expressions denoting emotional actions, states and processes. For those items missing from the preset dictionary, I manually supplemented semantic codes using existing entries as a basis. To identify emotions specifically associated with discussion of polyamory, I calculated the Dice coefficient (Rychlý 2008) of their frequency compared to the English-language section of the SIBOL corpus (Partington et al. 2022)

Discussion of emotion forms an important part of media reporting about polyamorous relationships and experiences. Highly salient terminology in the corpus focuses on experiences of desire (amorous, fetish, lust) and affection (lover, favorite, tlc ‘tender loving care’), but even more distinctively on those of unhappiness and conflict.

News reporting makes occasional space for the unique terminology developed by the polyamorous community to express its non-possessive relationship ideal (compersion, frubby), but much more frequently it dwells on the risk of jealousy and the experiences associated with it.

I argue that this persistent emphasis on the potential negative outcomes of polyamorous intimacy shows that the significant growth in polyamorous visibility over this period has not resulted in the reduction of the stigma against it.

References

- Adams, Megan Blythe and Nathan Rambukkana. (2018). "Why do I have to make a choice? Maybe the three of us could, uh...": Non-monogamy in videogame narratives. *Game Studies* 18(2). https://gamestudies.org/1802/articles/adams_rambukkana.
- Antalffy, Niko. (2011). Polyamory and the media. *Journal of Media Arts Culture* 8(1). http://scan.net.au/scan/journal/display.php?journal_id=157
- Barker, Meg-John. (2005). This is my partner and this is my ... partner's partner: Constructing a polyamorous identity in a monogamous world. *Journal of Constructivist Psychology* 18(1): 75-88. 10.1080/10720530590523107
- Barker, Meg-John and Darrem Langdridge. (2010). Whatever happened to non-monogamies? Critical reflections on recent research and theory. *Sexualities* 13(6): 748-772. 10.1177/1363460710384645
- Dow Jones & Company. (2024). *Factiva*. <https://global.factiva.com>
- Ferrer, Jorge N. (2018). Mononormativity, polypride and the "mono-poly wars". *Sexuality & Culture*, 22(3): 817-836. 10.1007/s12119-017-9494-y
- Hurson, Moira. (2016). *Networks of many loves: A history of alternative media in the polyamory movement*. [Ph. D. dissertation, University of Colorado at Boulder]. https://scholar.colorado.edu/concern/graduate_thesis_or_dissertations/pz50gw34r.
- Klesse, Christian. (2016). Marriage, law and polyamory. Rebutting mononormativity with sexual orientation discourse? *Oñati Socio-Legal Series* 6(6): 1348-1376.
- Moore Andrew and Paul Rayson. (2022). *PyMUSAS: Python multilingual UCREL semantic analysis system* (Version 0.3.0). Lancaster University. <https://ucrel.github.io/pymusas/>
- Moors, Amy C. (2017). Has the American public's interest in information related to relationships beyond "The Couple" increased over time? *The Journal of Sex Research* 54(6): 677-684. 10.1080/00224499.2016.1178208
- Moors, Amy C., Schechinger, Heath A., Balzarini, Rhonda and Sharon Flicker. (2021). Internalized consensual non-monogamy negativity and relationship quality among people engaged in polyamory, swinging and open relationships. *Archives of Sexual Behavior* 50(4): 1389-1400. 10.1007/s10508-020-01885-7
- Partington, Alan, Duguid, Alison, Marchi, Anna, Clark, Caroline, Taylor, Charlotte and John Morley. (2022). *SiBol: Corpus of English broadsheet newspapers 1993-2021*. Universities of Siena and Bologna. <https://www.sketchengine.eu/sibol-corpus/>
- Richardson, Leonard. (2023). *Beautiful soup documentation* (Version 4.12.1). <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
- Ritchie, Ani and Meg Barker. (2006). There aren't words for what we do or how we feel so we have to make them up: Constructing polyamorous languages in a culture of compulsory monogamy. *Sexualities* 9(5): 584-601. 10.1177/1363460706069987
- Rychlý, Pavel. (2008). A lexicographer-friendly association score. In *RASLAN 2008: Recent advances in Slavonic natural language processing*. P. Sojka and A. Horák (eds.) Masaryk University, 6-9. <https://nlp.fi.muni.cz/raslan/2008/papers/13.pdf>

Saxey, Esther. (2010). Non-monogamy and fiction. In *Understanding non-monogamies*. Meg Barker and Derren Langdrige (eds.) Routledge, 23–33.

Schippers, Mimi. (2016). *Beyond monogamy: Polyamory and the future of polyqueer sexualities*. New York University Press.

Séguin, Lea J. (2019). The good, the bad and the ugly: Lay attitudes and perceptions of polyamory. *Sexualities* 22(4): 669-690. 10.1177/1363460717713382

Semantic prosody evolution of the word "女权 [feminism/feminist]" in Chinese social media

A corpus-driven analysis

Siyuan Liu (Università di Bologna, Italy)

Keywords: semantic prosody, collocation, corpus, social media, feminism/feminist

Research on feminism/feminist vocabulary explores the usage, semantic prosody and social implications of terms related to feminism and gender equality. Semantic prosody refers to a word's tendency to attract words with similar semantic features, forming habitual collocations and a specific semantic atmosphere. This phenomenon extends beyond word boundaries, and this association is influenced by the user's native language but exhibits a certain breadth and stability. (Sinclair 1996: 87; Sinclair 2004a: 276; Bill 2000: 60; Partington 1998: 68) Examining the semantic prosody of a term provides intuitive insights into people's attitudes, emotions and, consequently, the cultural and social meanings behind it.

In China, the feminist movement was historically a longstanding yet niche voice. Since the #MeToo movement entered China in 2018, sparking a wave of complaints regarding sexual harassment and assault, feminism has gained unprecedented attention and discussion. However, concurrently, feminists face an increasing online "backlash," with the term "feminism/feminist" gradually acquiring negative connotations.

This study employs a data-driven approach, analyzing the evolution of feminist representation on China's prominent microblogging platform, Weibo. It investigates the semantic prosody evolution of the term "女权" (feminism/feminist), analyzing the attitudes and emotions of the online community, represented by social media users and the concept of "feminism/feminist". Semantic prosody is considered an extended meaning, where seemingly neutral items may acquire positive or negative semantic associations due to frequent co-occurrence with specific words.

The study collects original Weibo posts containing the term "女权" (feminism/feminist) from 2018 to 2023, constructing a balanced corpus with 128,046 posts. After processing, the total number of tokens is 6,692,258. Within a 4-4 window span, collocations with a minimum co-occurrence frequency of 20, minimum MI score of 3 and t-score of 2 are considered significant collocations (Xiao & McEnery 2006; Sinclair 1991).

The obtained significant collocations are categorized as positive, negative, or neutral. Ambiguous classifications are resolved by consulting dictionaries to ensure objectivity. Contextual factors, such as a word's usage in specific situations, are considered in meaning interpretation (Albertsson 2011: 8).

The analysis reveals a predominantly negative semantic prosody associated with the term "女权" (feminism/feminist) on Chinese social media, intensifying over the years. Despite feminism not being a new phenomenon, it appears that feminists and the feminist movement are still perceived as threatening, subjected to derogatory and

insulting language. Through such linguistic behavior, gender discriminatory language persists, with women remaining its primary targets (Albertsson 2011: 18).

References

- Albertsson, Jörgen. (2011). *'A man-hating lesbian feminist': A study of the collocations of the word feminist*.
- Louw, Bill. (2000). Contextual prosodic theory: Bringing semantic prosodies to life. In *Words in context, A tribute to John Sinclair on his retirement*. Heffer Chris & Saunt son Helen (eds.), 48-94.
- Partington, Alan. (1998). *Patterns and meanings*, 1-172.
- Sinclair, John. (1991). *Corpus, concordance, collocation*.
- Sinclair, John. (1996). *The search for units of meaning*, 1000-1032.
- Sinclair, John. (2004). New evidence, new priorities, new attitudes. In *How to use corpora in language teaching*, 271-299.
- Xiao, Richard and Tony McEnery. (2006). Collocation, semantic prosody and near synonymy: A cross-linguistic perspective. *Applied Linguistics* 27(1): 103-129.

Foregrounding and backgrounding through passive constructions

A corpus-based functional analysis of Chinese environmental discourse

Laura Locatelli (Università Ca' Foscari, Italy)

Keywords: Chinese linguistics, passive constructions, discourse functions, environmental discourse, corpus linguistics

Over the past decade, China has become a focal point of global concern, being recognized as one of the foremost contributors to greenhouse gas emissions and the consequential environmental issues (Leggett, Logan & Mackey 2019). Scholars have approached these challenges from various perspectives, with some delving into linguistic analyses of sustainable practices (Zhou 2016). However, limited attention, if any, has been directed towards exploring specific linguistic features employed in diverse environment-related contexts. Starting from the general assumption that the passive voice serves as a technique to obscure language, reflecting a way to avoid direct responsibility (Kahn 2006), this study investigates the role of the Chinese passive voice as a grounding device in discourse.

Adopting a functionally oriented approach, the three categories of passive constructions identified by Po-Ching & Rimmington (2016), i.e., marked passive constructions, notional passives and lexical passive structures, were isolated from the tailor-made multi-genre specialized *Corpus of Chinese Environmental Discourse* (CCED) and analyzed according to Hopper's foregrounding theory (1979). With the aim of discerning whether and which, passive constructions are used to highlight (foreground) portions of discourse or demote (background) others, a contrastive examination has been carried out focusing on the properties of foreground-background distinction: sequentiality of event, aspect, subject presupposition, predicate kinesis and realis. In addition, the results were further observed, both quantitatively and qualitatively, by intersecting the grounding system with the parameters of the transitivity analytical tool (Hopper & Thompson 1980).

Preliminary results suggest that Chinese passive constructions exhibit a low degree of participants' interactivity and transitivity. These structures tend to be linked to inanimate, abstract and vague entities, thereby contributing to a lower level of dramaticity in discourse. Furthermore, the passive voice is found to be commonly employed to introduce background information or downplay certain elements within a text, resulting in decreased processing priority and reduced accessibility. While distinctions among the three categories of passives can be discerned in terms of semantic prosody and emotional expressiveness, the passive voice in Chinese appears to serve as a means to convey an overall sense of indirectness, mitigate conflicts and promote social harmony in communication.

References

- Blühdorn, Hadarik. (2008). Subordination and coordination in syntax, semantics and discourse. 'subordination' versus 'coordination'. In *Sentence and Text. A Cross-linguistic Perspective*. Cathrine Fabricius-Hansen and Wiebke Ramm (eds.) Amsterdam: John Benjamins Publishing Company, 59-85.
- Hopper, Paul J. (1979). Aspect and foregrounding in discourse. In *Discourse and Syntax*. Talmy Givón (ed.) New York: Academic Press, 213-241.
- Hopper, Paul J. and Sandra A. Thompson. (1980). Transitivity in grammar and discourse. *Language* 56(2): 251-299.
- Kahn, Mary. (2006). The passive voice of science. In *Ecolinguistics reader: Language, ecology and environment*. Alwin Fill and Peter Muhlhausler (eds.) New York; London: Bloomsbury Publishing.
- Leggett, Jane A., Jeffrey Logan and Anna Mackey. (2019). *China's greenhouse gas emissions and mitigation policies*. Congressional Research Service.
- Li, Wendan. (2018). *Grounding in Chinese written narrative discourse 32*: Leiden: Brill.
- Po-Ching, Yip and Don Rimmington. (2016). *Chinese: A comprehensive grammar*. Oxon: Routledge.
- Zhou, S. (2016). Environmental, social and governance reporting in China. *Social and Environmental Accountability Journal* 36(1): 92-93.

Keywords of multiplicity

A corpus assisted discourse analysis of online multiplicity communities

Andrew Lustig (University of Toronto, Canada)

Gavin Brookes (Lancaster University, United Kingdom)

Mark McGlashan (Birmingham City University, United Kingdom)

Keywords: social media, discourse analysis, mental health, web-based health, complement

Multiplicity (also known as plurality) is a psychological phenomenon whereby a person experiences several selves within their single body. Historically, this was regarded as a disordered condition (American Psychiatric Association, 2004) but social media websites like Reddit have enabled people who experience multiplicity to gather online, discuss their experiences and contest the notion that multiplicity is necessarily disordered (Isler 2017). Previous work using corpus linguistics has studied a contested persecutory belief system known as gangstalking (Lustig et al 2021) and has showed that sufferers of a belief system labelled as pathological by mental health professionals can employ discursive strategies to construe a countervailing discourse wherein they contest the pathologization of their experiences.

This study examines language used in four of the largest multiplicity communities on Reddit to identify idiosyncratic language use in these communities. Of these four reddit communities, two identify multiplicity as pathological (r/DID, r/OSDD), one explicitly identifies it as nonpathological (r/Tulpas) and one contains a diversity of perspectives (r/plural). These different communities, united by the experience of having multiple selves but distinguished by their embrace or rejection of mental disorder provides an opportunity to examine how people engage with the discursive regimes of disordered and nodisordered multiplicity.

Our corpus of contains 800 threads and is comprised of the top 200 most upvoted threads for each of the four multiplicity communities described above (800 total threads).

To analyze this corpus, we employ a recently described methodology which combines key-keyword analysis with concepts from set theory to identify and compare keywords between corpora and find keywords that are used uniquely within each corpus (McGlashan & Krendel 2023). Key-key-words (Scott 1997) were identified by first generating keyword lists for each of the 200 threads for each of the four community subcorpora. We used the Webcorp Mini-Web reference corpus (<http://wse1.webcorp.org.uk/home/syn.html>) for keyword calculations. All keywords had to meet a Log-Likelihood threshold of ≥ 15.13 ($p < 0.0001$). Key-key-words were then identified for each subcorpus by aggregating the keyword lists for each of the 200 threads to rank keywords in terms of how often they are keywords in an individual thread. Key-key-word lists were then compared to find *complement keywords*, i.e. “key-key-words that [are] found to be isolated in their use to a single subcorpus” (McGlashan & Krendel 2023: 8).

Complement key-key-words within the r/DID (dissociative identity disorder) community focused on navigating challenges in interpersonal relationships. This includes romantic relationships (*partner*) as well as relationships with family (*mom, family*) and health care professionals (*misdiagnosed, professionals*). In addition, contributors focused on the challenges of receiving an accurate diagnosis of DID and the frustration of misdiagnosis. Keywords in the r/OSDD (Other specified dissociative disorder) community also focused on the challenges of receiving a diagnosis (*undiagnosed, psychologist*) and on concerns about being disbelieved (*impostor, pretending*). The r/plural keywords focused on the diversity of identities and experiences within the plural community (*neurodiversity*) and on concerns about including and excluding people from the community (*exclusionists, inclusion*). Finally, the r/Tulpa keyword lists focused on creative and imaginative processes.

Complement keywords in the disorder-oriented communities (r/DID and r/OSDD) focused on diagnosis, misdiagnosis, managing distress and promoting recovery and improved mental health. Both communities evinced concerns about being disbelieved and about the authenticity of experiences and diagnosis. The r/plural community focused on the genesis of plural experiences and symbolic social boundaries in the plural community. Finally, the r/Tulpa community foregrounds the creative, artistic and imaginative aspects of plurality.

References

- American Psychiatric Association, A. P. and American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders: DSM-IV*. Washington DC: American psychiatric association.
- Isler, Jacob J. (2017). Tulpas and mental health: A study of non-traumagenic plural experiences. *Research in psychology and behavioral sciences* 5(2): 36-44.
- Lustig andrew, Gavin Brookes and Daniel Hunt. (2021). Linguistic analysis of online communication about a novel persecutory belief system (gangstalking): Mixed methods study. *Journal of Medical Internet Research* 23(3): e25722.
- McGlashan, Mark and Alexandra Krendel. (2024). Keywords of the manosphere. *International Journal of Corpus Linguistics* 29(1): 87-115.
- Scott, Mike. (1997). PC analysis of key words — and key key words. *System* 25(2): 233-245.

A babel of voices

Making the Desert Island Discs corpus

Anna Marchi (Università di Bologna, Italy)

Adriano Ferraresi (Università di Bologna, Italy)

Keywords: corpus building, methodology, CADS

This paper presents a new corpus resource with wide-ranging potential in corpus-assisted discourse studies. We compiled a corpus containing over 2400 episodes (close to 12 million words) of *Desert Island Discs* transcripts, between 1951 and the present day and enriched it with fine-grained extra-linguistic information. In 80 years (the first broadcast was in 1942) the format of the iconic BBC radio program has remained unchanged: every week the host casts away on an imaginary desert island a well-known guest, with a choice of eight music tracks which guide the interview about the castaway's memories, personal life and career. This remarkable consistency makes it an interesting barometer of cultural British life over the past eight decades. But, as well as being a formidable dataset for studies of diachronic variation (Smith & Waters 2019), *Desert Island Discs* is also – in the words of one of the castaways – a “library of Babel of human stories, experience and language” (*Desert Island Discs*, David Mitchell 13 November 2020) and could be a versatile tool to explore a number of questions about identity, memory, emotions, storytelling and other typical interests of corpus-assisted discourse studies, in pursual of “non-obvious meanings” (Partington 2017). Because of its cultural relevance, *Desert Island Discs* has been object of several small-scale studies on presentation of the self (Cohen & Duberley 2013) and identity building. Lamont and Loveday, for example, “found that music was chosen to reflect guests’ national identity, gender identity and sometimes more hidden psychological states” (2020: 7). In order to allow for more systematic, larger scale analyses along these lines, we mined Wikidata using Python ad hoc scripts and were able to retrieve a rich set of metadata about almost all speakers (over 98% out of nearly 2300 unique speakers), including their age, profession, nationality, education and gender. These metadata were indexed as corpus search criteria, so as to make it possible to connect linguistic features and socio-demographic parameters. The full paper will describe the procedure that we developed to build the corpus, starting from the automatic transcription of episodes, to the annotation pipeline and finally the indexing with the NoSketch Engine platform, which will allow public sharing of the corpus. We will conclude by suggesting examples of analyses based on linguistic (as well as non-linguistic) questions, thus hopefully demonstrating the width and depth of research interests that a carefully made and finely annotated corpus can accommodate, as well as addressing relevant methodological issues of data availability and re-usability.

References

Cohen, Laurie and Joanne Duberley. (2013). Desert island data. Using BBC Radio 4's Desert Island Discs an organisation and management research. In *Unconventional Methodology in Organization and Management Research*. Bryman, Alan and David A. Buchanan (eds.) Oxford: Oxford University Press, 27-44.

Lamont, Alexandra and Catherine Loveday. (2020). A new framework for understanding memories and preference for music. *Music & Science* 3(1): 1-14.

Partington, Alan. (2017). Varieties of non-obvious meaning in CL and CADS: From 'hindsight post-dictability' to sweet serendipity. *Corpora* 12(3): 339-367

Smith, Nicholas and Cathleen Waters. (2019). Variation and change in a specialised register. A comparison of random and social linguistic sampling outcomes in Desert Island Discs. *International Journal of Corpus Linguistics* 24(2): 169-201.

Cast in memory

The discursive construct on of personal nostalgia in BBC Desert Island Discs

Anna Marchi (Università di Bologna, Italy)

Craig Evans (Swansea University, United Kingdom)

Keywords: CADS, discursive construction of emotions, operationalisation, nostalgia

The aim of this paper is to investigate evidence of distinct features of personal nostalgic discourse and explore ways of identifying ‘non-obvious’ (Partington 2017) linguistic markers that, working together, may create that nostalgic meaning we are able to recognise between the lines. In other words, we are in search of a shared language of nostalgia, that is, language that signals the emotion (Bednarek 2008) and language that potentially triggers the emotion. These pragmatic functions may manifest through multifaceted renderings, the methodological interest of this paper is understanding how a corpus-assisted approach (Partington 2004; Partington et al. 2013) can contribute to systematically and comprehensively (in terms of precision and recall) identify the mechanisms that express and invite nostalgic feelings.

The territory of our analysis is the Desert Island Discs corpus, containing over 2400 transcripts (almost 12 million words) of episodes of the iconic BBC Radio Four programme between 1951 and the present day. *Desert Island Discs* (Brown et al. 2017) is prototypically nostalgic in multiple ways. There is the “soundtrack of a lifetime” format which is intrinsically about reminiscing: every week the host is ‘cast away’ on an imaginary desert island with a well-known guest, who chooses eight music tracks to guide the interview through their memories, personal life and career. There is the castaway’s anticipatory nostalgia for what they will miss once stranded. There are the host’s explicit attempts to build up an expectation for nostalgia in framing the guests’ music selection (*Are you choosing nostalgically or?*); and then the island itself as a symbol of nostalgia, with its promise of a return to a simple life.

References to time and memory emerge as consistent core properties of nostalgic discourse, they are however not distinctive properties, as Marchi puts it ‘all nostalgic narratives are “once upon a time” narratives, but not all “once upon a time” stories are nostalgic’ (2023: 202). Relying on collocation analysis and on the close reading of concordances we retrieve, dissect and classify corpus evidence and show that it is the combination and agglomeration of multiple features that creates the nostalgic effect. Among them we find: the presence of sensory and emotional language, positive evaluation, the accumulation of detail, the attribution of sharpness, vividness or uniqueness to the memory, or alternatively the immanence of the moment depicted. The expression of authenticity and self-continuity (resonating with research in psychology, Sedikides et al. 2004). Finally, we come to define discursive nostalgia as: painting a backdrop of “back then” and spending time in the memory.

References

- Bednarek, Monika. (2008). *Emotion talk across corpora*. Basingstoke; New York: Palgrave MacMillan.
- Brown, Julie, Cook, Nicholas and Stephen Cottrell. (2017). *Defining the discographic self: Desert Island Discs in context*. Oxford: Oxford University Press.
- Marchi, Anna. (2023). Get back! Methods for exploring discourses of nostalgia and nostalgic discourses using corpora, *Elephant & Castle* 31(3): 199-218.
- Partington, Alan. (2004). Corpora and discourse, a most congruous beast. In *Corpora and Discourse*. John Morley and Louann Haarman (eds.) Bern: Peter Lang, 11–20.
- Partington, Alan. (2017). Varieties of non-obvious meaning in CL and CADS: From 'hindsight post-dictability' to sweet serendipity. *Corpora* 12(3): 339-367.
- Partington, Alan, Duguid, Alison and Charlotte Taylor. (2013). *Patterns and meanings in discourse: Theory and practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins.
- Sedikides, Constantine, Wildschut, Tim and Denise Baden. (2004). Nostalgia: Conceptual issues and existential functions. In *Handbook of experimental existential psychology*. Pyszczynski, Tom, Jeff Greenberg and Sander L. Koole (eds.) New York: Guilford, 200-214.

Mirror, mirror on the discursive wall, who's the most critical of them all?

A corpus-based study of the discourses of discourse analysis, with a special emphasis on the difference between critical and 'non-critical' approaches

Georg Marko (Universität Graz, Austria)

Keywords: critical discourse analysis, 'non-critical' discourse analysis, corpus-based discourse analysis, keyword analysis, semantic profile

Critical Discourse Analysis (CDA) is concerned with “linguistic and semiotic aspects of social processes and problems” (Fairclough, Mulderrig & Wodak 2011: 368). It was initially launched as an alternative to the insistence on unbiasedness, objectivity and abstraction from sociocultural contexts associated with traditional, ‘non-critical’ approaches to the study of discourse (including Conversation Analysis and sociolinguistic, pragmatic and psycholinguistic strands). However, since the time of its inception, Critical Discourse Analysis has developed into its own ‘brand’ (Billig 2000), so that not claiming the label *critical* today is sometimes (mis)interpreted as a disinterest in the social dimensions of discourse, including the construction of ideologies, social identities and relationships, the exercise of power and discrimination. But are there really significant differences between critical and ‘non-critical’ approaches to discourse analysis and to what extent do they adopt different scientific thought-styles (Pahta & Taavitsainen 2010: 23). These are the research questions this paper seeks to answer.

As CDA assumes that thought-styles are constructed in the way language is used in particular discourses, it seems plausible to apply its methods to itself and its ‘non-critical’ counterparts and investigate the language of discourse analysis. I will particularly look at conceptual strategies variously mentioned by critics as being too prominent ((a.), (b.), (e.) or not prominent enough ((c.), (d.)) in CDA (Dillon 1982; Stubbs 1997; Widdowson 1998; O’Halloran 2003):

- a) **Subjectification:** foregrounding the author’s beliefs and evaluations.
- b) **Moralization:** foregrounding evaluations and normativity.
- c) **Metatheorization:** foregrounding data and scientific practices of research.
- d) **Linguistic theorization:** foregrounding linguistic concepts, theories and models.
- e) **Extralinguistic contextualization:** foregrounding social, political and historical contexts.

I will use a corpus-based discourse analytical approach (e.g. Baker et al. 2008, Mautner 2009), examining two self-compiled 2-million-word corpora, each containing 300 articles from the last 15 years. The articles for the ‘critical’ corpus are taken from the journals *Discourse & Society*, *Social Semiotics*, *Journal of Language & Politics*, *CADAAD Journal* and *Critical Discourse Studies*, those for the ‘non-critical’ corpus from *Text & Talk*, *Discourse Processes*, *Discourse Studies*, *Discourse & Communication* and *Discourse, Context & Media*.

The corpora have been part-of-speech tagged and semantically annotated via Wmatrix5 (Rayson 2009) and are analysed with Wordsmith 9.

Methodologically, I partly proceed inductively, using keyword lists comparing the two corpora to see which semantic domains are quantitatively predominant and thus which aspects are more salient in critical and ‘non-critical’ approaches,¹ and partly deductively, focusing on particular constructions that can be argued to be associated with one of the aforementioned conceptual strategies.

The study is work in progress at this point but will be completed in the next three months. Although I have used the corpora for exploratory research (see Table 1), no major systematic results can as yet be reported.

References

- Baker, Paul, Gabrielatos, Costas, Khosravinik, Majid, Krzyżanowski, Michał, McEnery, Tony and Ruth Wodak. (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): 273-305.
- Billig, Michael. (2000). Towards a critique of the critical. *Discourse & Society* 11(3): 191-192.
- Dillon, George L. (1982). Whorfian stylistics. *Journal of Literary Semantics* 11(2): 73-77.
- Fairclough, Norman, Mulderrig, Jane and Ruth Wodak (2011). Critical Discourse Analysis. In *Discourse Studies. A Multidisciplinary Introduction* (2). Teun van Dijk (ed.) London: Sage, 357-378.
- Mautner, Gerlinde. (2009). Checks and balances: How corpus linguistics can contribute to CDA. In *Methods of Critical Discourse Analysis* (2). Ruth Wodak and Michael Meyer (eds.) Los Angeles: Sage, 122-143.
- O'Halloran, Kieran (2003). *Critical Discourse Analysis and language cognition*. Edinburgh: Edinburgh UP.
- Pahta, Päivi and Irma Taavitsainen (2010). Introduction. In *Early Modern English medical texts*. Taavitsainen, Irma and Päivi Pahta (eds.) Amsterdam: John Benjamins, 1-7.
- Rayson, Paul (2009). Wmatrix: a web-based corpus processing environment. Computing department, Lancaster University. <http://ucrel.lancs.ac.uk/wmatrix/> (Accessed 15 Jan 2024).
- Stubbs, Michael. (1997). Whorf's children: Critical comments on Critical Discourse Analysis. In *Evolving Models of Language*. Ann Ryan and Alison Wray (eds.) Clevedon: Multilingual Matters, 100-116.
- Widdowson, Henry (1998). The theory and practice of Critical Discourse Analysis. *Applied Linguistics* 19(1): 136-151.

¹ For a preview, see Table 1, which presents semantic profiles of the most important keywords, i.e. words occurring significantly more often in one corpus than in the other. The data seems to support the assumption that CDA focuses more on extralinguistic contextualization, but significantly less so, comparatively speaking, on metatheorization and linguistic theorization.

	Critical approaches	Non-critical approaches
Scientific practice General	<i>critique; textbook</i>	<i>assessment; cue; effect; condition; congruent/ congruency; experiment; eye-tracking; feedback; inconsistency; interference; judgment; null; pause, predict/predictive/prediction/predictor; priming; probe; random; rate/rating; screen; task; test; uncertainty; valence; variable; variance</i>
Quantification		<i>0.2; 0.5; 2; 3; 6; 7; 9; duration; measure; N400; SD; T1; T2</i>
Linguistic theory General	<i>CDS; DHA/discourse-historical; semiotics</i>	
Structure	<i>sign; transitivity</i>	<i>antecedent; NP; prosodic; sentence; referent;</i>
Text & communication	<i>logos; topos; multilingualism</i>	<i>coherence; conversational; epistemics; filler; formulation, hedge; interactional; maxims; metadiscourse; polite; abstract; chat; closing; CMC; comics; e-mails; expository; meme; movie; narrativity</i>
Cognitive processes & performance		<i>activation; comprehension; inference; learning; memory; non-native; problem-solving; processing; recall; skill</i>
Healthcare		<i>client; counsellor; diagnostic; patient; therapy/ therapist; deaf</i>
Nationality & geography & religion & languages	<i>BIH; borders; Bosnia/Bosnian; China; citizenship; city; Czech; Eastern; EU; Europe; Greece; Hindu; homeland; Islamic; Israel/Israeli; Jerusalem; Jews; land; Muslim; nation; Northern; Palestine/Palestinian; Serbia; transnational</i>	<i>Hungarian; Japanese</i>
Politics	<i>coalition; communist; conservative; election; elite; far-right; flag; governance; government; governmentality; hegemony; liberal; Nazi; parliament; patriot; populism; referendum; regime; minister; populist; liberty; Bush; Hitler; Thatcher</i>	

Society General	<i>ideological; logics; multiculturalism; societal; socio- political</i>	
Identity	<i>gay; homosexual/homosexuality; lesbian; LGBTQ; queer; same- sex; sex</i>	
Conflict	<i>sectarian/sectarianism; camp; terrorism/terrorist; victory; war</i>	
Inequality	<i>homeless; homophobia; inequality; nationalism; racism; slave/slavery; surveillance; colonial/colonialism</i>	
Economy	<i>brand/branding; capitalist/capitalism; tourism; trade</i>	

Table 1: Semantic profiles of the 100 most prominent keywords in the ‘critical’ corpus and the ‘non-critical’ corpus (only those expressions assigned to one of the semantic categories are included).

Climate change discourse in Polish newspapers over seven years

A corpus-based usage fluctuation analysis

Dagmara Mateja (Uniwersytet Warszawski, Poland)

Przemysław Szyc (Uniwersytet Warszawski, Poland)

Zuzanna Fałkowska (Uniwersytet Warszawski, Poland)

Tomasz Malinowski (Uniwersytet Warszawski, Poland)

Keywords: climate change, collocation, Usage Fluctuation Analysis, historical discourse, corpus linguistics

This research examines the development of climate change discourse in the Polish press from 2015 to 2022. The main objective is to analyse how media narratives in Poland have evolved during this period, which is marked by significant environmental and political events. This study also provides a methodological contribution by combining corpus-based discourse analysis (Baker 2006; Gabrielatos & Baker 2008; Jaworska & Krishnamurthy 2012; Baker et al. 2013; Wang 2018; Liu & Huang 2022) with the novel technique of usage fluctuation analysis (UFA) (Baker et al. 2017; Brezina 2018; McEnery et al. 2019, 2021; Gillings et al. 2023). The UFA technique allows us to identify statistically significant shifts in word usage over a period of time by tracking the fluctuation in word usage manifested by collocation. Using a combination of Usage Fluctuation Analysis (UFA) and corpus-based discourse analysis, the study examines the frequency and context of the terms *climate change* and *global warming* in a targeted corpus of Polish media.

For this study, we created a specialized corpus to investigate discourse on climate change. The corpus comprises articles published in the Polish press from 1 January 2015 to 31 December 2022. The corpus includes articles from the following daily newspapers: „Rzeczpospolita” and „Gazeta Polska Codziennie” and weekly newspapers: „Newsweek”, „Polityka”, „Wprost”, „Do Rzeczy” and „Sieci”. The corpus consists of 11,493 texts, with a total of 9,788,321 tokens. We used advanced methodology that includes machine learning algorithms to analyze these large data sets. Through this approach, patterns and themes related to climate change were identified without explicitly mentioning these terms, resulting in a broader range of relevant content.

Our analysis results confidently indicate that the media's presentation of climate change issues has undergone significant changes. This demonstrates a growing awareness and recognition of the importance of climate change in the media debate. Initially, climate issues received limited coverage, often overshadowed by other national concerns, but over time there has been a remarkable increase in volume and depth. The research clearly indicates a gradual shift in the Polish media's approach to climate issues, moving from minimal engagement and skepticism to a more proactive and informed discussion. Moreover, this study is a significant methodological contribution as it employs Usage Fluctuation Analysis (UFA) on a specialized corpus, providing new and valuable insights into the evolving dynamics of climate change discourse in the Polish context.

References

- Baker, Helen, Brezina, Vaclav and Tony McEnery. (2017). Ireland in British parliamentary debates 1803-2005: Plotting changes in discourse in a large volume of time-series corpus data. In *Exploring future paths for historical sociolinguistics*. Tanja Säily, Anita Auer, Minna Palander-Collin and Arja Nurmi (eds.) John Benjamins, 83-107.
- Baker, Paul. (2006). *Using corpora in discourse analysis*. Continuum.
- Baker, Paul, Gabrielatos, Costas and Tony McEnery. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British press*. Cambridge University Press.
- Brezina, Vaclav. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press.
- Gabrielatos, Costas and Paul Baker. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press 1996-2005. *Journal of English Linguistics* 36(1): 5-38.
- Gillings, Mathew and Carmen Dayrell. (2023). Climate change in the UK press: Examining discourse fluctuation over time. *Applied Linguistics* 45(1): 111-133.
- Jaworska, Sylvia and Ramesh Krishnamurthy. (2012). On the F word: A corpus-based analysis of the media representation of feminism in British and German press discourse, 1990-2009. *Discourse & Society* 23(4): 401-431.
- Liu, Ming and Jingyi Huang. (2022). "Climate change" vs. "global warming": A corpus-assisted discourse analysis of two popular terms in The New York Times. *Journal of World Languages* 8(1): 34-55.
- McEnery, Tony, Brezina, Vaclav and Helen Baker. (2019). Usage fluctuation analysis: A new way of analyzing shifts in historical discourse. *International Journal of Corpus Linguistics* 24(4): 413-444.
- McEnery, Tony, Baker, Helen, Brezina, Vaclav, Čermáková, Anna, Egan, Thomas, Hasselgård, Hilde and Sylvi Rørvik. (2021). Slavery and Britain in the 19th century. In *Time in Languages, Languages in Time*. Anna Čermáková, Sylvi Rørvik, Hilde Hasselgård and Thomas Egan (eds.) John Benjamins, 9-38.
- Wang, Guofeng. (2018). A corpus-assisted critical discourse analysis of news reporting on China's air pollution in the official Chinese English-language press. *Discourse & Communication* 12(6): 645-662.

Persuasion in COVID-19 political discourse beyond cognitive linguistics

Points of contact between sensorimotor concepts and morphosyntactic structures

Paul Mayr (Universität Innsbruck, Austria)

Christine Konecny (Universität Innsbruck, Austria)

Keywords: political discourse, persuasion strategies, cognitive linguistics, sensorimotor concepts, morphosyntax

In view of its strong impact not only on extralinguistic reality, but also on social behaviour and to some extent even on language use and system, the COVID-19 pandemic has attracted the interest of linguistic research since its beginnings: the various contributions on this topic, which have been published since the middle of 2020, range from those focusing on lexical innovations (Elia 2020; Sgroi 2020) and new models of word formation (Sgroi 2020; Pietrini 2021) to those anchored in the field of communication science as well as political linguistics in the broader sense (Koch & Thörle 2021; Wodak 2021).

The genre of public political speech itself had already been studied before the pandemic from numerous perspectives, including the morphosyntactic dimension (Danler 2007, 2020; Amenta/Strudsholm 2014). As far as political speeches during the COVID-19 pandemic are concerned, they have already been examined, too. So far, however, these studies focus on specific aspects and approaches, belonging above all to the field of cognitive linguistics, such as frame semantics (Wodak 2021) and conceptual metaphor theory (see in particular the several publications on the WAR metaphor, like e.g. Piazza 2020; Salvati & Verdighi 2020; Castro Seixas 2021; Semino 2021). On the other hand, there seems to be a lack of more specific analyses concerning e.g. sensorimotor concepts (Mayr 2023), which can in turn reflect conceptual metaphors or metonymies, or even going beyond (merely) cognitive linguistics, such as those focusing on morphosyntactic forms and structures that can be identified as characteristic of COVID-19 political discourse.

In this mainly qualitative study, a corpus of 10 selected speeches pronounced by Italian and Spanish politicians between 2020 and 2022 is analysed with regard to the use of sensorimotor concepts on the one hand (see the frequent use of lexical items such as It. *prendere, far fronte, lotta, di cuore* or Sp. *hacer frente, tomar, entender, batalla* etc.) and the use of recurrent morphosyntactic patterns on the other; among the latter, particularly salient seem to be different structures serving to express deontic modality, progressive periphrases, parenthetical clauses, accumulations of qualitative adjectives etc., as well as specific argument structures connected with highlighting or hiding certain semantic roles. In this way, we will demonstrate the persuasive functions and effects achieved through the interplay between both of these linguistic means and, more generally, between lexicon and morphosyntax.

References

- Amenta, Luisa and Erling Strudsholm. (2014). L'espressione della modalità deontica nei discorsi politici. In: *Dal manoscritto al web. Canali e modalità di trasmissione dell'italiano; tecniche, materiali e usi nella storia della lingua. Atti del XII congresso SILFI (Helsinki, 18-20 giugno 2012)*. Enrico Garavelli (eds.) Firenze, Cesati, 369-380.
- Castro Seixas, Eunice. (2021). War metaphors in political communication on Covid-19. *Frontiers in Sociology* 5. 10.3389/fsoc.2020.583680.
- Danler, Paul (2007). *Valenz und diskursive Strategien. Die politische Rede in der Romania zwischen 1938 und 1945: Franco – Mussolini – Pétain – Salazar*. Tübingen: Narr.
- Danler, Paul. (2020). *Der klassische Populismus Lateinamerikas. Politolinguistische Perspektiven auf Argentinien, Brasilien und Mexiko*. Bielefeld: transcript.
- Elia, Antonella. (2020). *Voci da una pandemia. Le parole della paura e della speranza*. Roma. Aracne.
- Koch, Christian and Britta Thörle. (2021). Vermessung und Vermittlung der Krise in den frühen Lageberichten zur Corona Pandemie: ein deutsch-französischer Vergleich. In *Zeitschrift für Literaturwissenschaft und Linguistik*. 51(3): 529-559.
- Mayr, Paul. (2023). Sensomotorische Konzepte in politischen Diskursen aus kontrastiv-linguistischer Perspektive: Beobachtungen zur Übersetzung sensomotorischer Konzepte vom Spanischen ins Deutsche, Französische und Italienische. In *SKOPOS – Revista Internacional de Traducción e Interpretación*. 13: 3-30.
- Piazza, Francesca. (2020). Metafore di guerra e guerra alle metafore. Sull'uso del lessico militare per parlare della pandemia di Covid-19. In *DNA – Di Nulla Academia, Rivista di studi camporesiani*. 2(2): 87-96.
- Pietrini, Daniela. (2021). *La lingua infetta. L'italiano della pandemia*. Roma. Treccani.
- Salvati, Lorena and Mirko Verdigi. (2020). La «Guerra al Covid-19» in occidente: il dilagare di una metafora. In *L'EPIDEMIA. Le parole e l'interpretazione delle parole*, Francesca Malagnini (eds.) Firenze. Cesati, 85-103.
- Semino, Elena. (2021). "Not soldiers but fire-fighters" – Metaphors and covid-19. *Health Communication* 36(1): 50-58.
- Sgroi, Salvatore. (2020). *Dal coronavirus al Covid-19. Storia di un lessico virale*. Alessandria: Ed. dell'Orso.
- Wodak, Ruth. (2021). Crisis communication and crisis management during COVID-19. *Global Discourse* 11(3): 329-353.

Text mining on mining

Investigating a sustainability report corpus through the lens of climate change

Anna Meyer (Montanuniversität Leoben, Austria)

Keywords: discourse analysis, corporate discourse, collocation networks, mining companies, sustainability reporting

Climate change is associated with human activities, which contribute to global warming. In recent years, publishing sustainability reports to account for operational impacts has become a trend. This can also be observed in the mining sector, which is naturally impact intensive in its operations. Significant greenhouse gas emissions stem from energyintensive mining and the processing. This paper makes a contribution to the field of public corporate discourse in relation to environmental impacts and investigates “corporate discourse and the outside world” (Jaworska 2020) using annual sustainability reports.

These voluntary reports provide a stage for self-representation, being composed to achieve this goal. Nonetheless, as Jaworska (2020) states, as writers and editors make choices, their favour of some over other choices offers valuable insight into how issues are presented to fit the company ideology. In the context of mining Poole (2016) has investigated a mining company’s discourse and position on environmental concerns.

The present paper contributes to the field by investigating the research question: *How does the text type of sustainability reports frame the words “climate” and “climate change”?* The hypothesis that the company uses the terms in a specific way and thus frames connected themes is tested via a comparison to a reference corpus. In order to investigate the research question a corpus of VALE’s – a global mining company - sustainability reports was created, spanning over eight years 2013-2020 and including 236.427 words. The VALE sustainability report corpus was searched for the terms *climate* and *climate change* using the software LancsBox. The focus was on identifying collocations of *climate* and the fixed expression *climate change* and to investigate their use in terms of frequency, position in collocation networks and directionality of association to frequent collocations. The results were compared to a reference corpus, a sample of the Corpus of Contemporary American English (COCA). The COCA sample was chosen due to its availability of relevant data, which encompassed the timespan of the texts in the VALE corpus (the sample COCA includes the years 1990-2019). The COCA samples consist of 8,9 million words with a word range between 962.652 and 1.345.414 words per text type (academic, blogs, fiction, magazine, newspaper, TV and movies subtitles and web pages).

Differences between the corpora are reflected in the themes associated with climate change. However, similarities exist on themes such as *risk* and *opportunity* and *energy*, which are prevalent in both corpora and are predominantly linked to discourse on climate change. The value of the analysis is especially enhanced by investigating

collocation networks (Brezina 2015), which provided interesting results on discourses on *risk* and *climate change* in the VALE sustainability report corpus.

Due to the difference in frequent collocations in each network, as could be shown impressively, the corpora create two fundamentally different networks. Another benefit of using collocation networks was the finding of asymmetric framings of the collocation *risk* and *opportunity / risks* and *opportunities* in the VALE corpus. Frequent mentions of *risk* are elaborated on and at times paired with opportunities, which never appears alone and is not described in detail. This gives rise to the argument that the frequent collocation *risks*, in the context of *climate change*, is softened by the use of (not elaborated) *opportunities*. This framing provides a glimpse into the softening narrative mining companies use in their reports to account for their environmental impact.

References

- Breeze, Ruth. (2013). *Corporate discourse*. London: Bloomsbury.
- Brezina, Vaclav, McEnery, Tony and Stephen Wattam. (2015). Collocations in context. A new perspective on collocation networks. *International Journal of Corpus Linguistics* 20(2): 139-173.
- COCA sample (2023) Linear text. Available at: <https://www.corpusdata.org/formats.asp> (Accessed 15 November 2023)
- Fairclough, Norman. (2008) *Analysing discourse textual analysis for social research*. London: Routledge.
- Gries, Stefan Th. (2013). 50-something years of work on collocations: What is or should be next. *International Journal of Corpus Linguistics* 18(1): 137-166.
- Heuer, Katrine. (2017). *The competitive advantage of sustainability reporting*. Baltimore. Johns Hopkins University.
- Jaworska, Sylvia. (2020). Corporate discourse. In *The Cambridge Handbook of Discourse Studies*. De Fina, Anna and Alexandra Georgakopoulou (eds.) Cambridge University Press, 666-686.
- Kolk, Ans. (2004). A decade of sustainability reporting: developments and significance. *International Journal of Environment and Sustainable Development* 3(1): 51-64.
- Linde, Charlotte. (2009). *Working the past: Narrative and institutional memory*. Oxford University Press.
- Poole, Robert. (2016). A corpus-aided ecological discourse analysis of the Rosemont Copper Mine debate of Arizona, USA. *Discourse & Communication* 10(6): 576-595.
- Wiechmann, Daniel. (2008). On the computation of collocation strength: Testing measures of association as expressions of lexical bias. *Corpus Linguistics and Linguistic Theory* 4(2): 253-290.

Gendered characters and gendered worlds in best-selling, contemporary Young Adult fiction

Sarah Jane Mukherjee (The Open University, United Kingdom)

Sally Hunt (The Open University, United Kingdom)

Maria Leedham (The Open University, United Kingdom)

Keywords: Young Adult fiction, gender, focus groups, librarian interviews, student voices

Children's and Young Adult (YA) literature offers 'windows and mirrors' onto the world (Bishop 1990), worldviews and experiences (windows) and self-reflections (mirrors). This paper focuses on YA fiction, which provides resources for identity building, as well as acting as a guide in navigating the world. Previous corpus linguistics research (Hunt 2015, in press) has shown highly gendered representations, as indicated through the linguistic patterning, of characters' appearance and agency in best-selling children's literature series (including *The Famous Five*, by Enid Blyton; *The Chronicles of Narnia* by CS Lewis). However, there is a paucity of work on representations of characters in recent young adult fiction using a CADS perspective. Therefore, little is known about the identity resources made available for young people in the literature they read as construed by the hidden discourse patterns of the language choices. This paper aims to address this gap and identify the discourses underlying the language patterns in relation to appearance and agency.

This presentation shares a CADS investigation of the 'windows and mirrors' on gender representations of the characters and the worlds constructed by female and male authors. This work draws on a newly compiled corpus of the 50 most commercially successful books for young adults (11-18 years) in the UK over a 5-year period (2017-2022).

Our analysis involves the investigation of impact of the role of gender in linguistic patterns in a unique corpus of c.5 million words. These findings were shared with and enriched by the insights of secondary school students and their school librarians, in focus groups and interviews.

The paper will report mainly on the following research questions:

1. What gendered patterns in characters' agency and appearance are constructed in YAF?
2. What differences and overlaps exist in the fictional worlds created by female and male authors?

To answer the first question, we searched for collocates of the gendered nouns *girl*, *woman*, *boy* and *man* and extracted the attributive modifiers of appearance preceding them. Those with the highest LogDice scores were investigated more fully in concordance lines for discourse patterns.

Female-authored (n=34) and male-authored books (n=16) were analysed separately through Wmatrix semantic tagging to explore what makes each subcorpus distinctive.

RQ1: Findings reveal interesting correlations between characters' gender and physical appearance, as shown in the following collocates:

- girls: *little, pretty, young, beautiful, baby*
- boys: *little, old, young, tall, invisible*
- women: *young, old, beautiful, blonde, elderly*
- men: *young, old, tall, dead, big*

Insights around gendered agency will also be explored in the presentation.

RQ2: Female and male authors constructed starkly distinct worlds in terms of the prominent semantic categories e.g. *emotions* and *warfare* respectively.

Findings were shared with young people and school librarians and their responses will be drawn on in the presentation.

References

Bishop, Rudine Sims. (1990). Windows and mirrors: Children's books and parallel cultures, In *Celebrating Literacy: Proceedings of the Annual Reading Conference at California State University*. Atwell, Margaret and Adria Klein (eds.)

Hunt, Sally. (2015). Representations of gender and agency in the Harry Potter series. In *Corpora and Discourse Studies: Integrating Discourse and Corpora*. Paul Baker and Tony McEnery (eds.) London. Palgrave Macmillan UK, 266-284

Hunt, Sally (in press). *Representations of genders in children's literature*. Palgrave Macmillan.

Metaphors of sustainability

A case study on media discourse

Laura Narisano (Libera Università di Bolzano, Italy)

Maria Christina Gatti (Libera Università di Bolzano, Italy)

Keywords: metaphor, metaphor identification, corpus linguistics, sustainability

Since the establishment of Conceptual Metaphor Theory (CMT, Lakoff & Johnson 1980) and of metaphor research, one of the major developments has been a shift from a reliance on researchers' intuition and *ad hoc* examples towards a focus on naturally occurring linguistic data contained in corpora, in order to gain a more comprehensive understanding of the characteristics of metaphorical language usage. One of the methods to locate metaphors in corpora requires selecting certain keywords and analysing their concordance lines (Deignan 2009). To determine the keywords in question, an area of investigation can be established *a priori* and treated as a potential source or target domain and lexicon related to it is considered. This provides insights either into the multivalency of the source domain, that is, the target domains it is used to conceptualise, or into the diversification of the target domain, that is, the various source domains used to conceptualise it (Goatly 1997).

In the present study, sustainability was considered as a target domain, with the aim of investigating the use of metaphors in the discourse on this topic. Sustainability, or sustainable development, understood as the ability of meeting the needs of the present without compromising the ability of future generations to meet their own needs (WCED 1987), has become increasingly relevant in recent years in public discourse and yet remains difficult to delineate clearly, due to its complexity (Purvis et al. 2018). When dealing with a complicated system of relatively unfamiliar concepts such as the notion of sustainability, metaphoric framings can provide a useful tool for its comprehension by prompting receivers to interpret the target issue using their knowledge of the source domain (Malt & Majid 2023). Since no single metaphor can provide a comprehensive representation of sustainability, different metaphorical mappings can be used to frame and offer different perspectives on the topic (Carew & Mitchell 2006), contributing to its diversification.

Considering the relevance of sustainability as not only a scientific or environmental issue, but also a social and political issue to be debated in the public sphere (Atanasova 2019), the manner in which it is discussed in newspapers becomes particularly relevant, due to the role they play in shaping public opinion (Baker et al. 2013) and, more specifically, in helping readers understand what sustainability is and how it can be achieved (Ziemann 2011). Consequently, a comparable corpus was created with articles from daily Italian and British newspapers mentioning, respectively, *sustainability* or *sustainable development* and *sostenibilità* or *sviluppo sostenibile*, starting from May 2022, when the 8th Environment Action Programme entered into force (European Commission

2022). Concordance lines for *sustainab-/sostenib-*, the keywords selected for the target domain, were analysed and figurative meanings were determined adapting the Metaphor Identification Procedure (Pragglejaz Group 2007, Steen et al. 2010). The results yielded by this approach demonstrate the role of metaphors in shaping the public discourse on sustainability, while contributing to the understanding of its conceptualisation.

References

- Atanasova, Dimitrinka. (2019). Moving society to a sustainable future: The framing of sustainability in a constructive media outlet. *Environmental Communication* 13(5): 700-711. 10.1080/17524032.2019.1583262.
- Baker, Paul, Gabrielatos, Costas and Tony McEnery. (2013). *Discourse analysis and media attitudes: The representation of Islam in the British Press*. Cambridge: Cambridge University Press.
- Carew, Anna L. and Cynthia A. Mitchell. (2006). Metaphors used by some engineering academics in Australia for understanding and explaining sustainability. *Environmental Education Research* 12(2): 217-231. 10.1080/13504620600690795.
- Deignan, Alice. (2009). Searching for metaphorical pattern in corpora. In *Contemporary Corpus Linguistics*. Paul Baker (ed.) London. Continuum International Publishing Group, 9-31.
- European Commission. (2023). Environment action programme to 2030. https://environment.ec.europa.eu/strategy/environment-action-programme-2030_en (Accessed January 27, 2023).
- Goatly Andrew. (1997). *The language of metaphors*. London. Routledge.
- Lakoff, George and Mark Johnson. (1980). *Metaphors we live by*. Chicago. University of Chicago Press.
- Malt, Barbara C. and Asifa Majid. (2023). Conceptual foundations of sustainability. *Topics in Cognitive Science* 15(3): 334-356. 10.1111/tops.12678
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1): 1-39.
- Purvis, Ben, Yong, Mao and Darren Robinson. (2018). Three pillars of sustainability: In search of conceptual origins. *Sustainability Science* 14(3): 681-695. 10.1007/s11625-018-0627-5
- Steen, Gerard J., Dorst, Aletta G., Herrmann, Berenike J., Kaal, Anna A. Krennmayr, Tina and Tryn tje Pasma. (2012). *A method for linguistic metaphor identification. From MIP to MIPVU*. Vrije Universiteit Amsterdam. Amsterdam: John Benjamins.
- Ziemann Andreas. (2012). Communication theory and sustainability discourse. In *Sustainability communication*. Jasmin Godemann and Gerd Michelsen (eds.) Houten: Springer, 89-96.
- WCED. (1987). *Our common future*. Oxford: Oxford University Press.

Representation of *politics* and *politician/s* in world Englishes

A GloWbE-based discourse analysis

Richmond Sadick Ngula (University of Botswana, Botswana)

Keywords: discourse, representations, world Englishes, variations, corpus-assisted discourse studies (CADS)

Research at the level of discourse in world Englishes has mainly focused on features of discourse as language in use (Kachru & Smith 2008; Mesthrie & Bhatt 2008, Kirkpatrick 2021) rather than on the analysis of 'discourse' as social practice (cf. Fairclough 2013; Fairclough, Mulderrig & Wodak 2011), one of whose key aims is to explore how people, social actors, or other phenomena are represented (or constructed) in reality. The lack of such studies may be due to, until recently, the non-availability of big data for world Englishes, which is required to enable researchers convincingly analyse a concept like *representation*. With the advent of big (textual) data in the form of a corpus, such as the 1.9-billion-word GloWbE corpus (Davies & Fuchs 2015), it is now possible to study discourse in this way across different varieties of English. In this paper, I explored how the related words *politics* and *politician/s* are constructed in varieties of native and non-native Englishes. I chose to search these words for analysis because of their crucial role in the events that shape how a society is organised.

Using a social theory of discourse (Foucault 1972; Burr 1995; Gee 2018), I analysed these two words in three inner circle varieties (USA, Great Britain and Australia) and three outer circle varieties (Ghana, Nigeria and Sri Lanka) in the GloWbE. The texts in the corpus are mainly from online sources, including informal blogs and newspapers. The analysis used the corpus methods of normalised frequencies, collocates and concordance lines (e.g., Baker 2023) and the discourse-historical approach to CDA (Wodak 2001). Quantitative differences were based on normalised frequencies as corpus sizes for the countries of focus varied.

Key findings include, first, that people in the outer circle countries generally tend to talk more about politics and politicians than those in the inner circle countries. Second, the analysis of both *politics* and *politician/s* highlights differences (and similarities) of discourse representation in not just between the inner and outer circle varieties but also between individual varieties. For instance, the analysis shows that while people in the inner circle countries discuss politics in terms of how it relates to or collaborates with other fields and in terms of preferable political ideologies, those in the outer circle countries tend to discuss politics more in terms of regional or national identities as well as in terms of political actors and institutions. A common representation of politics is that it is negatively evaluated in both the inner and outer circle varieties. In addition, between individual varieties, similarities and differences are observed in the way politicians are represented, especially in the discourses of *negative evaluation* and *nationality and external reference*. For instance, people in all the six countries consider politicians to be corrupt,

although this evaluation is discussed more prominently in the non-native countries. The findings suggest that representations of politics and politicians predominantly highlight culturally relevant variations between inner and outer circle countries and in specific countries, which may be attributed to differences in people's experiences of (and attitudes towards) political activities and actors in their geographical contexts.

References

- Baker, Paul. (2023). *Using corpora for discourse analysis* (2). London: Bloomsbury.
- Burr, Vivien. (1995). *An introduction to social constructionism*. London: Routledge.
- Davies, Mark and Robert Fuchs. (2015). Expanding horizons in the study of world Englishes with the 1.9 billion word Global Web-based English Corpus (GloWbE). *English World-Wide* 36(1): 1-28.
- Fairclough, Norman, Mulderrig, Jane and Ruth Wodak. (2011). Critical discourse analysis. In *Discourse Studies. A multidisciplinary Introduction*. Teun A. Van Dijk (ed.) Sage, 357-378.
- Fairclough, Norman. (2013). *Critical discourse analysis: The critical study of language* (2). London. Routledge.
- Foucault, Michel. (1972). *The archaeology of knowledge*. London. Tavistock.
- Gee, James Paul. (2018). Introducing discourse analysis: From grammar to society. In *Cultures, contexts and world Englishes*. Kachru, Yamuna and Larry E. Smith (eds.) London: Routledge.
- Kirkpatrick andy (2021). *Routledge handbook of world Englishes* (2). London: Routledge.
- Mesthrie, Rajend and Rakesh M. Bhatt. (2008). *World Englishes: The study of new linguistic varieties*. Cambridge. Cambridge University Press.
- Wodak, Ruth. (2001). The discourse-historical approach. In *Methods of critical discourse analysis*, Ruth Wodak and Michael Meyer. London: Sage, 63-94.

How undercover is terminological variation in supranational law? A corpus-assisted exploration of police incitement

Jekaterina Nikitina (Università degli Studi di Milano Statale, Italy)

Keywords: terminological variation, human rights discourse, legal terminology, legal genres, police incitement

This study explores the terminological landscape surrounding policing and undercover operations within supranational law, with a specific focus on the European Court of Human Rights (ECtHR). The proper use of legal terminology has been recognized as a crucial factor contributing to the clarity and consistency of the ECtHR's operation (Brannan 2023). As the ECtHR deals with cases originating in 46 jurisdictions of the Council of Europe (CoE), it is exposed to differing conceptualizations, linguistic realizations and legal definitions and implications of the practice when a law enforcement agent induces a person to commit a crime. This exposure results in a significant amount of terminological variation (Freixa 2002) in the pinnacle genre within the system of genres (Bazerman 1994) at the ECtHR: the judgment. Consequently, potential friction points and harmonization challenges may arise in the application of supranational jurisprudence within domestic legal landscapes.

The study discerns where and how the ECtHR's terminology diverges from domestic interpretations, shedding light on the nuanced complexities that emerge at the crossroads of national and supranational legal systems.

The materials comprise a corpus of judgments in English dealing with police entrapment cases (Art. 6 of the European Convention on Human Rights) along with other procedural genres within the ECtHR system of genres. The main methodological framework of the (cross-linguistic) corpus-assisted discourse study (Gillings, Mautner & Baker 2023) is supplemented by insights from legal terminology (Biel & Kockaert 2023) and legal translation studies (Cheng et al. 2014; Biel et al. 2019). The study builds on the classification of legal terms at international organizations by Prieto Ramos (2014) and, specifically, at the ECtHR by Brannan (2013, 2018, 2023) and Peruzzo (2019), supplementing them with phraseological perspectives where applicable, as the terms under analysis pertain to the category of the so-called multiword terms, combining elements of terminology and phraseology (Nikitina 2019: 272).

The findings identify the co-existence of diverging terminological choices across different cases and, surprisingly, even within the procedural documents dealing with the same case (application, case communication, written pleadings, judgment). In applying the Convention, the ECtHR relies on the so-called generic terms or terms that underwent "autonomous interpretation" (Peruzzo 2019: 81), but it also operates with terms coming from domestic law of a given CoE State and/or other instruments of international law (Brannan 2023). Multiword terms such as "agent provocateur", "undercover operation", "police incitement" and "(police) entrapment" coexist with system-bound terminology

(Peruzzo 2019; Nikitina 2022), i.e., terms rooted in the national legal system that must remain marked as national terms (e.g., “investigative test”, “operative experiment”). The latter appear in the overview of domestic laws and facts and, for some languages, can be marked by the presence of translation couplets (Šarčević 1985: 131). A significant amount of hidden or covert translation from non-official languages carried out by the ECtHR lawyers at the case communication stage, significantly preceding judgment drafting, is found to be a factor contributing to terminological variation, along with overt translation across the two official languages of the Court, English and French, at the judgment stage.

The findings thus offer a critical reflection on the potential challenges resulting in the misuse or incorrect terminological choices, underscoring the dynamic interplay between supranational and domestic legal discourses.

References

- Bazerman, Charles. (1994). Systems of genres and the enhancement of social intentions. In *Genre and new rhetoric*. Aviva Freedman and Peter Medway (eds.) London: Taylor and Francis, 79-101.
- Biel, Łucja and Hendrik Kockaert. (2023). *Handbook of terminology (3): Legal terminology*. Amsterdam: John Benjamins.
- Biel, Łucja, Engberg, Jan, Rosario, Ruano, Martín and Vilelmini Sosoni. (2019). *Research methods in legal translation and interpreting: Crossing methodological boundaries*. London: Routledge.
- Brannan, James. (2013). Coming to terms with the supranational: Translating for the European Court of Human Rights. *International Journal for the Semiotics of Law* 26: 909-925.
- Brannan, James. (2018). Specificities of translation at the European court of human rights: Policy and practice. In *Institutional translation for international governance: Enhancing quality in multilingual legal communication*. Fernando Prieto Ramos (ed.) London: Bloomsbury, 170-180.
- Brannan, James. (2023). From the domestic to the supranational. The terminology of “expulsion” as used at the European court of human rights. In *Handbook of terminology (3): Legal terminology*. Łucja Biel and Hendrik Kockaert (eds.), 327–350. 10.1075/hot.3.dom2.
- Le Cheng, King Kui Sin and Anne Wagner. (2014). *The Ashgate handbook of legal translation*. Farnham. Ashgate.
- Freixa Aymerich, Judit (2002). La variació terminològica. Anàlisi de la variació denominativa en textos de diferent grau d'especialització de l'àrea de medi ambient. In *Filologia Catalana*. Barcelona. Universitat de Barcelona.
- Gillings, Mathew, Mautner, Gerlinde and Paul Baker. (2023). *Corpus-assisted discourse studies (Elements in corpus linguistics)*. Cambridge: Cambridge University Press.
- Nikitina, Jekaterina. (2019). The discourse of bioethics in the ECtHR case-law: A corpus-based perspective. *Lingue e Linguaggi* 29: 269-289.
- Nikitina, Jekaterina. (2022). The discourse of the European court of human rights: The role of case communications in the introduction of system-bound elements. In *Language for international communication: Linking interdisciplinary perspectives: Language for specific purposes in the era of multilingualism and technologies (4)*. Laimdota Ločmele (ed.) Riga: University of Latvia Press, 97-106. 10.22364/lincs.2023.08
- Peruzzo, Katia. (2019). *National law in supranational case-law: a linguistic analysis of European court of human rights judgments in English*. Trieste: EUT Edizioni.
- Ramos, Fernando Prieto. (2014). Parameters for problem-solving in legal translation: Implications for legal lexicography and institutional terminology management. In *The Ashgate Handbook of Legal Translation*. Le Cheng, King Kui Sin and Anne Wagner (eds.) Farnham: Ashgate, 121-134
- Šarčević, Susan. (1985). Translation of culture-bound terms in laws. *Multilingua* 4(3): 127-134.

RQTR to test relevance in query terms to build a bilingual corpus of media representation of Ukrainian refugees

Aroa Orrequia-Barrea (Universidad de Cádiz, Spain)

Keywords: corpus linguistics, corpus building, critical discourse analysis, media representation, Ukrainian refugees

On 24th February 2022, Russia invaded Ukraine and a war broke out. Since then, millions of Ukrainians have been forced to leave their houses and fled to different parts of the European continent.

This research is part of a larger project which looks into media representation of Ukrainians in the UK and Spanish press and compare it with the discursive construction of other migrant groups in the press (Baker & McEnery 2005; Gabrielatos & Baker 2008; Baker et al. 2008; KhosraviNik 2009, 2010a, 2010b).

The aim of this study is to show the results of employing the *relative query term relevance* (RQTR, henceforth) technique proposed by Gabrielatos (2007) to identify the best candidates for a query to build a bilingual corpus.

Five steps were followed to carry out the study (Gabrielatos 2007: 31). First, two pilot corpora, one in each language, were compiled over the period of a year with the terms *refugee* and *asylum* as seeds (Gabrielatos 2007: 12). We downloaded the news articles published in the third week of each month, since February 2022 until February 2023, to reduce the presence of seasonal words. Second, keywords were extracted to see the candidate terms that can be potentially included in the query. For this purpose, we used the written and news subcorpora of the BNC2014 from #Lancsbox X (Brezina & Platt 2023) for the English corpus. For the Spanish corpus, it was compared with the news and European Spanish domain subcorpora of the esTenTen from *Sketch Engine* (Kilgariff & Renau 2013). In this case, we also needed to calculate log likelihood (Rayson, n.d.). Third, we also included some introspective candidates that were not retrieved by the keywords but were semantically related to the topic. The fourth step was to calculate the *query term relevance* score (QTR, henceforth) to obtain the baseline for the subsequent steps. Fifth, the RQTR is calculated using the QTR and the baseline. Finally, RQTRn was calculated so that the results were normalised.

The RQTR procedure deemed relevant many candidates, but the researcher's subjectivity is required to choose which ones can be included in the query. In our case, we took the candidates that were key and positive in RQTR, namely the highest result, 100. The query for the English corpus resulted in: *refugee** OR *asylum** OR *immigra** OR *migra** OR *evacuee** OR *evacuat** OR *sanctuary* OR *exodus* OR *flee** OR *fled* AND *Ukraine** OR *Ukrainian**. In the case of the Spanish, the query was slightly different: *refugiad** OR *asilo* OR *inmigr** OR *migr** OR *evacu** OR *exodo* OR *huir** OR *huy** OR *acog** AND *Ucrania* OR *ucranian** OR *ucrani**

The RQTR procedure provides researchers with a useful mechanism to choose scientifically relevant terms for a query avoiding the presence of non-related articles or noise. Additionally, it allows researchers to consider whether the introspective ones are actually relevant. RQTR is also a helpful way to start the analysis as it also points to research areas that can be interesting for the future analysis (Gabrielatos 2017).

References

- Baker, Paul and Tony McEney. (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, Paul, Gabrielatos, Costas, Khosravinik, Majid, Krzyżanowski, Michał, McEney, Tony and Ruth Wodak.(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.
- Brezina, Vaclav and William Platt. (2023). *#LancsBox X* [software], Lancaster University, <http://lancsbox.lancs.ac.uk>.
- Gabrielatos, Costas. (2007). Selecting query terms to build a specialised corpus from a restricted-access database. *ICAME journal* 31(5): 5-44.
- Gabrielatos, Costas and Paul Baker. (2008). Fleeing, sneaking, flooding: A corpus analysis of discursive constructions of refugees and asylum seekers in the UK press, 1996-2005. *Journal of English linguistics* 36(1): 5-38.
- KhosraviNik, Majid. (2009). The representation of refugees, asylum seekers and immigrants in British newspapers during the Balkan conflict (1999) and the British general election (2005). *Discourse & Society* 20(4): 477-498.
- KhosraviNik, Majid. (2012). 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, Adam and Irene Renau. (2013). esTenTen, a vast web corpus of Peninsular and American Spanish. *Procedia-Social and Behavioral Sciences* 95: 12-19.
- KhosraviNik, Majid. (2010). Actor descriptions, action attributions and argumentation: Towards a systematization of CDA analytical categories in the representation of social groups. *Critical discourse studies* 7(1): 55-72.
- Rayson, Paul. Log likelihood and effect size calculator. <https://ucrel.lancs.ac.uk/llwizard.html> (Accessed January 8, 2024).

'Homeless' in the cost-of-living crisis

A corpus-assisted analysis of British national media coverage

Tamsin Parnell (University of Nottingham, United Kingdom)

Keywords: homelessness, cost-of-living, crisis, media representations, corpus-assisted critical discourse analysis

The language used to describe homeless people can pigeonhole individuals negatively and have implications for social responses to homelessness (Daly 1996). People tend to deal with people experiencing homelessness by distancing themselves to minimise fear, guilt, or shame (Daly 1996). This is never more evident than in the discursive divide between "us" and "them" or the dehumanising collectivisation of the noun phrase "the homeless". Through the third person pronoun 'they', people experiencing homelessness 'become an amorphous, remote, alien mass lacking individuality or even humanity' (Daly 1996: 8). In other words, language can become a 'tool of manipulation' (Daly 1996: 9). This is especially problematic in the case of media accounts of homelessness, which are a key source of information for many people (Schneider, Chamberlain & Hodgetts 2010; Power 1999). It is more concerning given that 'even when members of the general public are already familiar with homelessness in their own communities, media help them make sense of their experiences and prescribed strategies for their responses to the homeless' (Hodgetts, Cullen & Radley 2005).

This paper constitutes a corpus-assisted critical discourse analysis exploring representations of people experiencing homelessness during the beginning of the cost-of-living crisis in the left-leaning national newspaper, *The Guardian* and the right-leaning online news outlet *MailOnline*. The two specialised corpora consist of 405,860 words and 446,715 words, respectively. Using Lancsbox X, the paper examines collocations of the noun 'homeless' in both corpora to uncover the most prominent ways in which people experiencing homelessness are framed in the two newspapers. Through a close linguistic analysis of textual examples, it uncovers the linguistic strategies through which people experiencing homelessness are constructed. It considers the discursive consequences of these representations during a period of economic crisis and in light of statistics that represent homeless populations in the UK. It also contemplates whether the moment of the cost-of-living crisis represents an opportunity for increased empathy in the UK, based on the idea that people are less likely to blame people for their own poverty during periods of economic crisis (Van de Mieroop 2011).

The paper demonstrates that while there are some shared collocates between the two corpora (e.g., 'people' and 'population'), *The Guardian's* collocates focus on people on the brink of homelessness ('made', 'becoming', 'risk') and help for those experiencing homelessness ('shelter', 'charity', 'shelters', 'charities'), whereas *MailOnline's* collocates include words related to crime ('shot'), addiction ('drug', 'addicts'), males ('man', 'men') and generally negative words ('problem'). The paper considers the question of *who* is

constructed as homeless in each case and the degree to which these representations reflect official statistics. It questions the implications of these representations for empathy for people experiencing homelessness among readers, which has been found to be lower in the UK than in other countries.

References

- Daly, Gerald. (1996). *Homeless: Policies, strategies and lives on the Streets*. Oxford: Routledge.
- Hodgetts, Darrin, Cullen, Andrea and Alan Radley. (2005). Television characterizations of homeless people in the United Kingdom. *Analyses of Social Issues and Public Policy* 5(1): 29-48. 10.1111/j.1530-2415.2005.00054.x
- Power, Gerald. (1999). Media image and the culture of homelessness: Possibilities for identification. In *Reading the homeless: The media's image of homeless culture*. Eungjun Min (ed.) London: Praeger, 65–83.
- Schneider, Barbara, Chamberlain, Kerry and Darrin Hodgetts. (2010). Representations of homelessness in four Canadian newspapers: Regulation, control and social order. *Journal of Sociology and Social Welfare* 37(4): 147-172.
- Van De Mieroop, Dorien. (2011). Identity negotiations in narrative accounts about poverty. *Discourse & Society* 22(5): 565-591. 10.1075/ni.21.1.03van

A corpus-driven comparative study of the complexity of patient information leaflets and their effectiveness in British and Italian domains

Nicola Pelizzari (University of Portsmouth, United Kingdom)

Keywords: patient information leaflets, lay-friendliness, corpus linguistics, readability

Patient information leaflets (PILs) serve as users' sources of legitimate information on medicine administration, safety and side effects. Far from being mere reference material, PILs guide consumers to make informed healthcare decisions. This means that PILs become controversial when they do not deliver comprehensive information and can thus be a source of confusion to patients (Askehave & Zethsen 2014). As such, PILs should balance two complementary and opposing requirements. On the one hand, they must meet the needs of user-friendliness for a diastatically heterogeneous audience, on the other hand, they must comply with scientific accuracy and a set of stringent criteria established by current regulations (Zethsen 2022). Nevertheless, expectations on PILs are often disregarded as the results seem to fall short of the intended lay-friendliness targets (Fage-Butler 2013; Okoro 2022). PILs carry out an indirect and unidirectional form of communication (from the sender to the addressee), which does not allow the addressee, in case of doubt, to obtain immediate clarification and, on the contrary, delegates the task of providing explanations to a third interlocutor, external to the communicative interaction, namely the doctor or pharmacist. Moreover, it is a communication deferred in place and time in which the recipient plays a primarily passive role. With this in mind, it is understandable how the concepts of medication adherence and compliance can be undermined if users have not understood what a statement instructs them to do (Albert, Bix & Bridgeman 2014). Although the European Union and the United Kingdom have implemented rather homogeneous regulations for PILs, there is a high degree of variation in how the approach to readability is addressed in different national and linguistic domains (Miglietta 2012). This article presents the methodology and preliminary results of my PhD research. This contribution aims to identify features that hinder patients' accessibility to PILs by using a corpus-driven approach (Biber 2015) to collect quantitative data on four different corpora that respectively comprise the PILs of the top-selling over-the-counter and prescription drugs in the UK and Italian contexts. This comparative work sheds light on the stylistic, syntactic and lexical differences in PILs. It shows that these features can vary the level of accessibility to PILs and require future investigations in a universally recognised field, with a view to improving those documents that directly affect all human beings and their fundamental right to health.

References

- Steven M. Albert, Bix, Laura, Bridgeman, Mary M., Carstensen, Laura L., Dyer-Chamberlain, Margaret, Neafsey, Patricia J. and Michael S. Wolf. (2014). Promoting safe and effective use of OTC medications: CHPA-GSA National Summit. *The Gerontologist* 54(6): 909-918.
- Askehave, Inger and Karen Korning Zethsen. (2014). A comparative analysis of the lay-friendliness of Danish EU patient information leaflets from 2000 to 2012. *Communication & Medicine* 11(3).
- Biber, Douglas. (2012). Corpus-based and corpus-driven analyses of language variation and use. In *The Oxford handbook of linguistic analysis*. Oxford University.
- Fage-Butler, Antoinette. (2013). Improving patient information leaflets: Developing and applying an evaluative model of patient-centredness for text. *Communication & Medicine* 10(2): 105.
- Miglietta, Annarita. (2012). Leggibilità dei bugiardini?: fra Italia e Inghilterra non c'è partita. In *Lingua italiana d'oggi*. 9: 163-192.
- Okoro, Roland N. (2022). Can your patient read this? The need to move to patient-centered medication product labels and patient information leaflets. *Journal of the American Pharmacists Association* 62(5): 1528-1530.
- Zethsen, Karen Korning. (2022). PILs that give you a headache? The lay-friendliness of non-pharmacy restricted, over-the-counter information leaflets. *Qualitative Health Communication* 1(1): 44-60.

Investigating ecological change and climate crisis through diachronic corpus-assisted ecolinguistics

Robert Poole (University of Alabama, United States of America)

Keywords: diachronic studies, ecolinguistics, climate crisis

In this presentation, I investigate how social data captured in large corpora can complement existing scientific evidence of climate change and serve as indicators of our worsening ecological crisis. Indeed, there is ample scientific data for our present ecological crisis, but we have largely left unexplored the social data present in large collections of language and how this data may also reflect the climate crisis and demonstrate our changing understandings of and relationships with the physical world. While it is broadly accepted that social and technological changes are reflected in language change, much less explored is the relationship between language change and ecological change.

Thus, the present study explores the shifting representations of the ecology-relevant words *wilderness* and *hurricane/s* in order to investigate the shifting perceptions, understandings and relationships with these entities and events. The approach integrates the techniques of discourse-oriented corpus studies with the framework of ecolinguistics, as it implements a data driven approach for investigating the role of language on “the life-sustaining interactions of humans, other species and the physical environment” (ecolinguistics-association.org).

In the presentation, I will briefly present case studies of two eco-keywords: *wilderness* and *hurricane/s*. For the former, I explore shifting collocational patterns with *wilderness* across the approximately two-hundred years of language use captured in the Corpus of Historical American English. For the latter, a diachronic collocation analysis of adjectives co-occurring with *hurricane/s* across the six five-year periods in the Corpus of Contemporary American English was conducted. In the analysis, collocation data were collected and the Kendall's Tau correlation coefficient for each collocate was calculated in order to assess the strength of diachronic change. For *wilderness*, the data demonstrates that framings of grandeur and vitality have declined while framings reflecting poor health have increased. For *hurricane/s*, the data indicate that adjectives indicating larger size, greater severity and increased strength are more frequently co-occurring with the node terms across the last three decades.

The presentation asserts the affordances of diachronic analysis for revealing changing perceptions and understandings of the physical world and climate change-related events. Finally, the presentation forwards a case for continued diachronic corpus studies in ecolinguistics to complement existing physical science data of climate change, to demonstrate how language use reflects beliefs and practices that contribute to the climate crisis, but also to explore how emergent language patterns align more positively with ecological wellbeing, justice and sustainability.

Half-witted women and exceptional men?

A corpus-assisted study of gendered social actors in two large heritage mountaineering corpora

Claudia Posch (University Innsbruck, Austria)

This presentation examines patterns of social actor representation and gender in two large cultural heritage corpora. The research presented compares the domain-specific *New Zealand Alpine Journal Corpus*, an 8-million-word corpus of the New Zealand Alpine Journal (NZAJ) and *Alpenwort*, a 19-million-word corpus of the Austrian Alpine Journal (ZAV). The two corpora are comparable in terms of the period they cover and their thematic orientation, namely mountains and mountaineering. Mountaineering texts have not received much attention in discourse studies, although, like other leisure and sport discourses, they reflect social and political developments. Even more than in leisure and sports discourses, what counts as a sporting achievement in mountaineering depends on discursive events, without spectators or referees present. As a result, there is an abundance of texts on the subject that are well suited to CADS.

Beyond their communicative function, mountaineering texts reflect certain conceptualisations, beliefs, values and norms and behaviours that are prevalent in the group of mountaineers. Moreover, the discursive products are consumed by a very wide audience, consisting of other professional mountaineers, amateurs, or people who only occasionally hike or climb and even "armchair alpinists" (Walter 1984). Historically, ideas about masculinity (and femininity vice versa) have been based on the idea of the British mountaineer or conqueror and this is true of both, the Austrian and the New Zealand alpinist traditions. In modern texts the focus has shifted to the athletic body and competition between nations. Men are still the dominant figures in these discourses (as well as in sport in general) (Apollo 2021; Moraldo 2020), despite the increasing participation of women. It is therefore not surprising that women are under-represented in the texts¹. When they are there, they are represented in significantly different ways than men are. The presentation will show how those ways are different by comparing frequently occurring patterns of key actor terms (Baker 2010, 2014; Taylor 2013) and their collocations by using collocation network graphs (Desagulier, 2020a; 2020b). To facilitate cross-linguistic comparison, van Leeuwen's (1996a) taxonomy of social actor representation was used to assign the types of nomination and categorisation (similar to Krendel et al. 2021) that social actor terms and their collocates entail. For example, is there a difference in the types of modifiers that can be used with different social actor terms?

¹ In the data, the representation of other types of (gendered) identities is not even traceable.

References

- Baker, Paul. (2010). Will ms ever be as frequent as mr? A corpus-based comparison of gendered terms across four diachronic corpora of British English. *Gender and Language* 4(1). 10.1558/genl.v4i1.125.
- Desagulier, Guillaume. (2020a). Plotting collocation networks with r and graph (blogpost). <https://corpling.hypotheses.org/3341>
- Desagulier, Guillaume. (2020b). Plotting collocation networks with r: 'hoard' vs. 'stockpile' in the coronavirus corpus (blogpost). <https://corpling.hypotheses.org/3300>
- Krendel, Alexandra, McGlashan, Mark and Veronika Koller. (2021). The representation of gendered social actors across five manosphere communities on reddit. *Corpora* (in press) 17(2):1-25.
- Taylor, Charlotte. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora* 8(1): 81-113. 10.3366/cor.2013.0035
- Van Leeuwen, Theo. (1996). The representation of social actors. In *Texts and practices: Readings in critical discourse analysis*. Normann Fairclough, Carmen Rosa Caldas-Coulthard and Malcolm Coulthard (eds.) London: Routledge, 32-70.
- Walter, J. A. (1984). Death as recreation: Armchair mountaineering. *Leisure Studies* 3(1): 67-76. 1080/02614368400390051

The use of adjective phrase inversion in learned and scientific exposition

Carlos Prado-Alonso (Universidad de Oviedo, Spain)

Keywords: adjective phrase inversion, scientific discourse, involvement, corpus, written English

In the last years, subject-dependent inversion — constructions in which the subject follows the entire verb phrase in a declarative clause, as in “On the near corner was Herb’s Gas Station” or “Upstairs was a bedroom and a bathroom”— has been the subject of extensive research from a corpus-based perspective (cf. De Wit 2016, Prado-Alonso 2016, Teixeira 2016, or Duffley 2018, among others).

On the basis of the kind of phrasal category occurring as clause-initial constituent, five different types of subject-dependent inversion have been traditionally distinguished in the literature on the topic: noun phrase, adverb phrase, verb phrase, adjective phrase and prepositional phrase inversion. Broadly speaking analyses of subject-dependent inversion carried within the functional paradigm have focused mainly on inversions following a locative constituent (as in the examples provided above) which subsumes “spatial locations, path and directions and their extension to some temporal and abstract locative domains” (Bresnan 1994: 75). In this sense, these studies have neglected the analysis of some inversion types, such as adjective phrase inversion --as in “Equally important is the state of affairs in Uganda” or “Conspicuous among the winners was Michael Jordan”-- which do not contain such locative clause-initial constituents.

In order to bridge the gap mentioned above, the present paper offers a corpus based analysis of the distribution and pragmatic function of adjective phrase inversion in Present-day English scientific texts taken from the Corpus of Contemporary American English (COCA); cf. Davies (2008). The scientific texts under analysis deal with natural sciences, medicine, technology and engineering, politics, law, education, humanities and social sciences. In the study, they have been grouped into two main categories: Scientific Exposition (natural sciences, medicine, technology and engineering) and Learned Exposition (Politics, Law, Education, Humanities and Social Sciences); see Biber (1998).

The distribution and frequency of adjective phrase inversions in the corpus texts is compared to Biber’s (1988) well-known multidimensional textual analysis. On the basis of a statistical analysis and a comparison, it is argued that the distribution of adjective phrase inversion in these scientific texts may allow the presence of the writer in the text in order to change the reader’s focus of attention. It will also be argued that the more involved a scientific text, the greater number of adjective phrase inversions may be expected and that the construction itself may be considered a discourse marker through which involvement is coded in scientific writing. Finally, it will also be demonstrated that because of its more involved nature, Learned Exposition makes a more frequent use of adjective phrase inversion than Scientific Exposition.

References

- Biber, Douglas. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bresnan, Joan. (1994). Locative inversion and the architecture of universal grammar. *Language* 70: 72-131.
- Davis, Mark. (2008). The corpus of contemporary American English (COCA). <https://corpus.byu.edu/coca/>
- De Wit, Astrid. (2016). The relation between aspect and inversion in English. *English Language and Linguistics* 20(1): 107-128.
- Duffley, Patrick. (2018). The cognitive structure of full-verb inversion and existential structures in English. *Cognitive Semantics* 4(2): 184-229.
- Prado-Alonso, Carlos. (2016). A constructional analysis of obligatory XVS structures. *Studia Anglica Posnaniensia* 51(1): 51-82.
- Teixeira, Joana. (2016). Locative inversion and stage topics: A cross-linguistic study. *Discours: Revue de linguistique. Psycholinguistique et Informatique* 19: 1-28.

The good, the tasty and the delicious

Evaluative strategies in online eating shows

Sofia Rüdiger (Universität Bayreuth, Germany)

Keywords: language and food, collocation, evaluation, identity-construction, YouTube

Eating shows have recently emerged as a new genre of performative food consumption in the virtual realm. Based originally on Korean live-streamed Mukbang (cf. Choe 2019), eating shows are performed asynchronously and usually feature a solo-performer eating copious amounts of food while talking to the non-co-present viewers (Rüdiger 2020). Since their conception in South Korea, eating shows and Mukbang have evolved into a truly global phenomenon, capturing the interest of a diverse viewership, contributing to the spread of Korean culture throughout the world and playing an important role in creating digital conviviality and eating communities.

In this study, I use a corpus of 100 YouTube eating shows (by 10 performers; ca. 250,000 words) to investigate how eating show hosts evaluate the food items that they consume. In particular, I draw on the notion of subjective vs. objective constructions in food assessment (Wiggins & Potter 2003), the use of gustatory exclamations (cf. Wiggins 2002) such as *mmm* (to indicate pleasant food experiences) and *ew* (to indicate unpleasant food experiences) and adjectives commonly used to describe taste (Bagli 2021). The corpus was analyzed using AntConc, drawing on the manual analysis of concordance lines for disambiguation (where necessary), to determine word frequencies and collocation patterns of the items and constructions of interest.

The analysis of the corpus shows that the performers prefer to use objective evaluations (e.g., *it is good*; relating to the object of the evaluation) over subjective evaluations (e.g., *I liked the burger*; marked by verbs indicating personal stance such as *love*, *hate*, *(dis)like*). In terms of gustatory vocabulary, the taste descriptors are predominantly positive (e.g., *delicious* n=0.25 per thousand words [ptw] vs. *bland* n=0.04ptw). Despite a range of culinary adjectives being in use in the shows to describe the food experience to the viewers (referring to gustatory, haptic, olfactory, visual and auditory aspects of the food), these are outnumbered by general adjectives to index a superficially positive stance (e.g., *good* n= 3.27ptw, *nice* n=0.88ptw vs. *tasty* n=0.05ptw; *flavorful* n=0.02ptw). It is thus also not surprising that positive gustatory exclamations prevail in the dataset (e.g., *mmm* n=4.07ptw vs. *eww* n=0.07ptw). The collocation analysis of *mmm* (5L/5R) reveals that positive gustatory experiences marked by this expression are also accompanied by exclamations of surprise (e.g., *woah*, *wow*) to indicate unexpected (but pleasant) flavors and general delight in the taste encountered. The positive taste experience is further underlined by frequent repetitions of *mmm* as well as its co-occurrence with lexical items indicating positive stance (cf. the occurrence of *yum*, *yummy*, *bomb(s)* and *awesome* in the Top 15 collocates of *mmm*). Altogether, food evaluations on eating shows are predominantly positive and the shows thus deviate from genuine ‘taste

testing' episodes. Emphasis lies on the creation of the identity of a lowbrow gourmand and an atmosphere of gustatory pleasure, allowing the performers to create digital eating communities around other activities which are present on the shows, such as story-telling.

References

- Bagli, Marco. (2021). *Tastes we live by: The linguistic conceptualisation of taste in English*. Berlin: De Gruyter.
- Choe, Hanwool. (2019). Eating together multimodally: Collaborative eating in mukbang, a Korean livestream of eating. *Language in Society* 48(2): 171-208.
- Rüdiger, Sofia. (2020). Dinner for one: The use of language in eating shows on YouTube. In *Talking About Food: The Social and the Global in Eating Communities*. Sofia Rüdiger and Susanne Mühleisen (eds.) Amsterdam; Philadelphia: John Benjamins, 145-165.
- Wiggins, Sally and Jonathan Potter. (2003). Attitudes and evaluative practices: Category vs. item and subjective vs. objective constructions in everyday food assessments. *British Journal of Social Psychology* 42: 513-531.
- Wiggins, Sally. (2002). Talking with your mouth full: Gustatory 'mmm's and the embodiment of pleasure. *Research on Language and Social Interaction* 35(3): 311-336.

A multifactorial analysis of CAN and MAY in essays written by American, German and Korean university students

Mi-Lim Ryoo (Korea Maritime & Ocean University, Republic of South Korea)

Keywords: learner corpus research, modal verbs, German English, Korean English, multifactorial analysis

In recent years, learner corpus research (LCR) has become an established subfield of general corpus linguistics utilizing more sophisticated statistical techniques (Gries & Deshors, 2020). From an LCR perspective, the present study compares the use of the nearly synonymous modal verbs, *can* and *may* by English learners from different L1 backgrounds (in this case, German and Korean) with native English uses, applying multifactorial methods which allow the analyst to identify linguistic features that distinguish non-native uses from native uses (Deshors, 2014). The study aims to find out which linguistic features contribute to nonnativeness and what relationship, if any, there is between these linguistic features and learners' L1.

The study hypothesized that both German and Korean learners of English would show a higher preference for *can* over *may* than native English speakers because *can* is generally more frequent (cf. Neff et al. 2003) and thus the default form in learners' English (Deshors & Gries 2014). This discrepancy would be greater for Korean learners than for German learners because the L1 of Korean learners lacks semantically equivalent forms of *can* and *may* (i.e., Korean uses the same expression (~ *eu*) *su itda*) for both *can* and *may*), whereas German has *können* and *dürfen*, which are more or less equivalent to English *can* and *may* respectively. Furthermore, *can* in Korean English would show the characteristic of amalgamating the morphosyntactic and semantic features of *can* and *may* in native English, whereas the linguistic behavior of *can* and *may* in German English would deviate relatively less from that used by native speakers.

The data for this study was derived from three comparable corpora: The American subsection of the Louvain Corpus of Native English Essays (LOCNESS-AE), the German subsection of the International Corpus of Learner English (ICLE-GE) and the Korean Learner English Corpus (KLEC). Following the Behavioral Profile approach which helps the analyst to examine the interactions between multiple variables and a linguistic choice (Gries 2012, 2018; Deshors 2014), a total of 2,140 occurrences of *can* and *may* in the three corpora were first manually annotated and analyzed.

It was found that among the three corpora, KE employed *can* most frequently (1,181 times (0.7%)), followed by GE (871 times (0,52%)) and then AE (720 times (0.43%)), whereas *may* appeared in exactly the opposite order (83, 147 and 236 times, respectively). The differences in the frequency of *can* and *may* between the corpora were statistically significant, except for between AE and GE for *can* ($p=.984$), which means that the frequency of *can* in GE was more similar to that in AE than KE was to AE. The overall results supported the initial hypothesis that GE and KE would show a higher preference for *can*

than AE and KE more than GE. The interactions between the co-occurring morphosyntactic and semantic features of *can* and *may* in native and learner English are currently being assessed using logistic regression modelling.

References

- Collins, Peter. (2009). *Modals and quasi-modals in English*. Rodopi.
- Depraetere, Ilse and Susan Reed. (2020). Mood and modality in English. In *The Handbook of English Linguistics*. Bas Aarts, April MS McMahon and Lars Hinrichs (eds.) John Wiley & Sons Lt, 207-227.
- Deshors, Sandra C. (2014). A case for a unified treatment of EFL and ESL: A multifactorial approach, *English World-Wide* 35(3): 279-307.
- Deshors, Sandra C. (2014). A case for the multifactorial assessment of learner language: The uses of *may* and *can* in French-English interlanguage. In *Polysemy and synonymy: corpus methods and applications in Cognitive Linguistics*. Justyna Robinson and Glynn Dylan (eds.) John Benjamins.
- Gries, Stefan Th. (2018). On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally. *Journal of Second Language Studies* 2(1): 276-308.
- Gries, Stefan Th and Sandra C. Deshors. (2014). Using regressions to explore deviations between corpus data and a standard/target: Two suggestions. *Corpora* 9(1): 109-136.
- Gries, Stefan Th and Dagmar Divjak. (2010). Quantitative approaches in usage-based cognitive semantics: Myths, erroneous assumptions and a proposal. In *Quantitative cognitive semantics: Corpus-driven approaches*. Dylan Glynn and Kerstin Fischer (eds.) Mouton de Gruyter, 333-354.
- Huddleston, Rodney and Geoffrey Pullum. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- McEnery, Tony and Nazareth Amselom Kifle. (2002). Epistemic modality in argumentative essays of second language writers. In *Academic Discourse*. Ken Hyland and John Flowerdew (eds.) Pearson Education, 182-195.
- Neff, JoAnne, Dafouz, Emma, Honesto Herrera, Martínez, Francisco, Juan Pedro Rica, Mercedes Díez, Prieto, Rosa and Carmen Sancho. (2003). Constructing learner corpora: The use of modal and reporting verbs in the expression of writer stance. In *Extending the Scope of Corpus-Based Research*. Granger, Sylviane and Stephanie Petch-Tyson (eds.) Rodopi, 212-230.
- Palmer, Frank Robert. (2001). *Modal and modality*. Cambridge University Press.

Misogyny vs. solidarity

Social media polarization over revenge porn

Marta Sánchez Cócera (Universidad Complutense de Madrid, Spain)

Keywords: revenge porn, CADS, social media discourse, cybercrimes, appraisal

The purpose of this paper is to explore polarization in the discourse of social media commentators over crimes of revenge porn. Revenge porn refers to “the nonconsensual sharing of intimate images” (Bond & Tyrrell 2021), videos, or other intimate material, which are ultimately weaponized against its victims. As is the case with other sorts of crimes, revenge porn cases are highly discussed on social media. Previous research on social media has demonstrated that users engage in polylogues that (a) reproduce prejudice and stigma or, contrarily, (b) resist unequal ideologies as a form of activism (Jones et al. 2022; Garcés-Conejos Blitvich & Bou-Franch 2019; Palomino-Manjón 2022). In light of this, this paper aims to explore the discursive construction of the phenomenon of Revenge Porn, its victims and its perpetrators on the social network platform YouTube. YouTube constitutes an interactive site, where polylogic communication and confrontation are maximized, as users can directly react to the content of videos, to other users, or to the general audience (Bou-Franch & Garcés-Conejos Blitvich 2016: 65). Considering this, the present study analyzes a corpus of 15,000 comments (ca. 440,000 words) from four different YouTube videos tackling Revenge Porn cases to reveal conflicting ideologies among online users. All the YouTube videos selected had a reportative style in which specific crimes were described and victims explained their experiences. Additionally, they were all published by officially verified journalistic and informative media.

The paper adopts a Corpus-Assisted Discourse Studies (CADS) approach (Baker & McEnery 2015; Gillings et al. 2023; Partington et al. 2013) and uses the software tool *Sketch Engine* to retrieve data that describes the most recurrent themes in the corpus. To this end, a keyness analysis was run, capturing the unique lexical patterns that characterize the corpus; the top 50 keywords were then analyzed in context. Additionally, due to the gendered nature of revenge porn crimes, the adjective and verb collocates of the lemmas GIRL, WOMAN, MAN, BOY as subject and object were explored using the ‘Word Sketch’ function to scrutinize the representation of victims and perpetrators. Finally, a more detailed analysis of evaluative resources was carried out drawing on Appraisal Theory (Martin & White 2005). Results show that a segment of commenters use victim-blaming discourse to construct female victims of revenge porn, ascribing traditional gendered sexual roles to them, while perpetrators are, in turn, exonerated from all responsibility. Other users, however, emphasize blame on both perpetrators and society for stigmatizing female victims, using positive appraisal of victims to build solidarity toward them.

References

- Baker, Paul and Tony McEnery. (2015). *Corpora and discourse studies: Integrating discourse and corpora*. Palgrave Macmillan.
- Bond, Emma and Katie Tyrrell. (2021). Understanding revenge pornography: A national survey of police officers and staff in England and Wales. *Journal of Interpersonal Violence* 36(5-6): 2166-2181. 10.1177/0886260518760011
- Bou-Franch, Patricia and Pilar Garcés-Conejos Blitvich. (2016). Gender ideology and social identity processes in online language aggression against women. In *Exploring Language Aggression against Women*. Bou-Franch, Patricia (ed.) John Benjamins, 59-81.
- Garcés-Conejos Blitvich, Pilar and Patricia Bou-Franch. (2019). Introduction to analyzing digital discourse: New insights and future directions. In *Analyzing digital discourse: New insights and future directions*. Pilar Garcés-Conejos Blitvich and Patricia Bou-Franch(eds.) Palgrave Macmillan, 3-22.
- Gillings, Mathew, Mautner, Gerlinde and Paul Baker. (2023). *Corpus-assisted discourse studies*. Cambridge University Press.
- Jones, Lucy, Małgorzata Chałupnik, Mackenzie, Jai and Louise Mullany. (2022). 'STFU and start listening to how scared we are': Resisting misogyny on Twitter via #NotAllMen. *Discourse, Context & Media* 47: 100596. 10.1016/j.dcm.2022.100596
- Martin, James R. and Peter R. White. (2005). *The language of evaluation: Appraisal in English*. Palgrave Macmillan.
- Palomino-Manjón, Patricia. (2022). Feminist activism on Twitter: The discursive construction of sexual violence and victim-survivors in #WhyIDidntReport. *Journal of Language Aggression and Conflict* 10(1): 140-168.
- Partington, Alan, Duguid, Alison and Charlotte Taylor. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. John Benjamins.

Metaphors we interpret by: What's with WATER?

Charlotte Taylor (University of Sussex, United Kingdom)

Keywords: metaphor, collocation, media discourse, CDS

Since the publication of Johnson & Lakoff's (1980) seminal work on conceptual metaphor, (critical) discourse studies has embraced metaphor as a tool for understanding how discourses are construed. With regard to newspaper discourse alone, recent years have seen studies of metaphor in relation to representation of migration (e.g. Catalano & Musolff 2019; Charteris-Black 2006), climate change (e.g. Cotton et al. 2019; Woods et al. 2012), healthcare (e.g. Cotter et al 2012; Semino 2021), international conflict (e.g. Palmero 2019; Sahlane 2013), gender (e.g. Koller 2004), homelessness (e.g. de Melo Resende 2020) and many more. Their relevance is strengthened yet further by experimental research in cognitive linguistics and metaphor studies which indicates metaphor use intensifies reader responses to texts (e.g. Hart 2018). Certain metaphor source domains (e.g. WAR or FIRE) are observed to recur across in studies of diverse topics in the media which raises questions about the strength of relationship between a target (e.g. immigrants) and source domain (e.g. WATER). These discourse studies have been supported by the rise of corpus linguistics but may have neglected one of the basic tenets of the field: that frequency of a linguistic phenomenon must be considered as a *relative* concept. We have empirical evidence regarding the frequency of metaphor in the press (16.4% of lexical items are metaphorically used according to Krennmayr 2015), but no measurement of which concepts are frequently drawn on to frame discussions of events, people or ideas. In one of the few contributions to address frequency of source domains, Karlberg & Buell (2005) found 17% of articles in *Time* magazine (1981–2000) contained at least one WAR metaphor. This suggests that the presence of WAR metaphors is a feature of the genre and this should be taken into account in interpreting the significance of such metaphors with reference to any specific target. In this paper, I report on a pilot study from a larger project which will establish this type of baseline for the British press. The pilot study focusses specifically on WATER metaphors in eight UK newspapers. In the first phase, the entire output from two days' reporting was manually annotated for WATER metaphors. In the second phase, corpus tools were used to scale this up to analysis of the entire output over a full year. This paper will summarise the frequency of WATER metaphors (by paper and article) and outline the most common targets (people, money, concepts etc.). The findings will allow researchers to assess whether the presence of WATER metaphors in their data is characteristic of framing of that particular group or phenomenon. The broader aim is to establish and evaluate a replicable frame to be repeated on the same data for different source domains and across other genres and languages.

References

- Catalano, Theresa and Andreas Musolff. (2019). "Taking the shackles off": Metaphor and metonymy of migrant children and border officials in the US. *metaphorik.de* 29.
- Charteris-Black, Jonathan. (2006). Britain as a container: Immigration metaphors in the 2005 election campaign. *Discourse & Society* 17(5): 563-581.
- Cotton, Matthew, Barkemeyer, Ralf, Renzi, Barbara Gabriella and Giulio Napolitano. (2019). Fracking and metaphor: Analysing newspaper discourse in the USA, Australia and the United Kingdom. *Ecological Economics* 166: 106426.
- Cotter, Colleen, Samos, Danniella and Deborah Swinglehurst. (2021). Framing obesity in public discourse: Representation through metaphor across text type. *Journal of Pragmatics* 174: 14-27.
- de Melo Resende, Viviane. (2020). A controversy in Folha de S. Paulo: Critical discourse reflections on the representation of homelessness and the coloniality of being. *Discourse & Society* 31(6): 584-606.
- Hart, Christopher. (2018). 'Riots engulfed the city': An experimental study investigating the legitimating effects of fire metaphors in discourses of disorder. *Discourse & Society* 29(3): 279-298.
- Karlberg, Michael and Leslie Buell. (2005). Deconstructing the 'war of all against all': The prevalence and implications of war metaphors and other adversarial news schema in TIME, Newsweek and Maclean's. *Journal of Peace and Conflict Studies* 12(1): 22-39.
- Koller, Veronika. (2004). *Metaphor and gender in business media discourse: A critical cognitive study*. Springer.
- Krennmayr, Tina. (2015). What corpus linguistics can tell us about metaphor use in newspaper texts. *Journalism Studies* 16(4): 530-546, 10.1080/1461670X.2014.937155
- Palmero, Celeste Moreno (2019). Nuclearism: Media discourse, image-schemata and the Cold War. *Critical approaches to discourse analysis across disciplines* 11(2):145-168.
- Sahlane, Ahmed. (2013). Metaphor as rhetoric: newspaper Op/Ed debate of the prelude to the 2003 Iraq War. *Critical Discourse Studies* 10(2): 154-171.
- Semino, Elena. (2021). "Not soldiers but fire-fighters"—Metaphors and covid-19. *Health Communication* 36(1): 50-58.
- Woods, Ruth, Fernández, Ana and Sharon Coen. (2012). The use of religious metaphors by UK newspapers to describe and denigrate climate change. *Public Understanding of Science* 21(3): 323-339.

Exploring discourse organization through metadiscourse, textual colligation, recursion and discourse patterning

Aleksandar Trklja (University Innsbruck, Austria)

Keywords: discourse organization, textual colligation, metadiscourse, recursion, corpus linguistics

The present paper introduces a novel data-driven methodological approach to the investigation of discourse organization. Although the approach applies to any kind of text, the focus of the present study is on judgments from the Court of Justice of the European Union (CJEU) and the United Kingdom Supreme Court (UKSC), extracted from EUCLCORP, a large multilingual corpus of legal judgments (AUTHOR). A peculiarity of legal judgments is the absence of systematic structural markings such as headings or paragraphs.

In the first part of the paper, the results of the analysis of formulaic expressions serving as metadiscursive devices (Hyland 2005) associated with specific textual colligation (Hoey, 2005) will be presented. The examination reveals that paragraph-initial formulaic metadiscursive signalling devices (PIMFE) indicate text development and repetition degree in both CJEU and UKSC judgments. CJEU judgments predominantly feature PIMFE indicating logico-deductive relations (e.g. It follows from this, It follows from the foregoing, It is clear from the) while UKSC judgments exhibit more variation, including Self-mentions, Frame markers and Importance markers. This analysis suggests that the discourse structure of CJEU judgments is based on a “lego-building block” principle, characterized by a high degree of automation and standardization in language patterns.

Building on the results presented in the first part, the second part of the paper discusses how PIMFE signals textual organization in terms of discourse grouping, utilizing Hoey's (1983) discourse patterns (e.g., Problem-Solution pattern, Opportunity-Taking Opportunity pattern). Notably, the paper introduces a methodological model that investigates the interaction between interactive metadiscursive devices (IMD), discourse pattern devices (DPD) and textual colligation. The results of the analysis demonstrate that when when IMD and DPD exhibit the property of textual colligation they signal discourse patterning. Thus, Frame markers and Frame markers and Evidentials indicate Question-Answer and Claim-Response patterns, respectively. Finally, the findings also show that the discourse organisation of legal judgments can be described in terms of nested relations between individual discourse patters. This is explained by borrowing the notion of linguistic recursion from syntax (Hauser et al. 2002) which has already been employed in conversation analysis (Levinson 2013). The results of the analysis show that in CJEU judgments the Claim-Response pattern is embedded within the Question-Answer pattern.

References

Hauser, Marc D., Chomsky, Noam and W. Tecumseh Fitch. (2002). The faculty of language: what is it, who has it and how did it evolve? *Science* 298: 1569-1579.

Hoey, Michael. (1983). *On the surface of discourse*. London. George Allen and Unwin.

Hoey, Michael. (2004). Textual colligation: a special kind of lexical priming. *Advances in corpus linguistics: Papers from ICAME, 23. Göteborg 22–26 May 2002*.

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

Levinson, Stephen C. (2013). Recursion in pragmatics. *Language* 89(1): 149-162.

Construction of the dichotomy of West versus East in the American and Russian newspapers

Corpus-driven discourse analysis of the news coverage of Ukrainian revolution 2014

Svitlana Tubaltseva (Richmond American University London, United Kingdom)

Keywords: antagonistic discourse, CADS, semantic prosody, collocations

The conflict between Ukraine and Russia has dominated the international agenda for the past two years, receiving extensive coverage worldwide. Yet, few have linked the ongoing warfare to the Ukrainian Revolution Euromaidan of 2014, which previously underscored internal divisions in the country. Most works on Euromaidan focus on the role of Ukrainian social media in mobilizing the protests (Bohdanova 2014; Heintz 2013; Kapliuk 2013; Lokot 2013; Onuch 2015; Knoblock 2020) rather than on a geopolitical positioning of Ukraine between two superpowers – Russian and the USA. This research paper delves into the construction of the dichotomy of 'Us and Them' or 'West and East' in the American and Russian press in the discourse on Euromaidan.

Given the complexity of the 'Us and Them' dichotomy as a socio-political and discursive phenomenon, an interdisciplinary approach is essential. The paper utilises the epistemological traditions of CADS research on warfare (Brym et al. 2014; Byun & Hollander 2015; Du 2016; Elena 2016; Lawless & Chen 2016; Lukin 2019) and complement it with the evaluative focus (Martin & White 2005; Partington 2015; Stubbs 2001) and the antagonistic discourse model (Carpentier 2018). While CADS approach offers tools to retrieve evaluative patterns and prosody from key collocations, the antagonistic model offers deeper insights into how the Us versus Them dichotomy is constructed and realized in discourse.

Regarding the data selection, two criteria, such as a large circulation and different political affiliation, are applied. Thus, two American newspapers, *The Washington Post* (right-wing) and *The New York Times* (left-wing) and two Russian newspapers *Izvestia* (pro-government) and *Kommersant* (pro-opposition) have been chosen. Due to the lack of existing corpora, a tailor-made corpus of each newspaper coverage during the Euromaidan is compiled.

The paper recognises the limitation of focusing on two newspapers from each region, however, the analysis still provides intriguing data which can shed light on the construction of the Us and Them dichotomy in each newspaper. Although all the newspapers construct a clear dichotomy of Us and Them, the structure of the dichotomy varies and appears to be more complex and multifaceted than previously discussed. Another significant finding reveals that despite a similar agenda, the newspapers from two contexts deploy different discursive and ideological techniques. The American newspapers invest more in the construction of the negative image of Russia and provide a more geopolitical framing of the event. The Russian newspapers, due to its geographical

and cultural proximity to Ukraine, offer a more personalized approach and focus on their role in the crisis rather than constructing the negative image of the West. This discursive discrepancy clearly highlights Euromaidan 2014 as a visible geopolitical event which was used by other countries to promote their ideological positions (Russian and the USA) and which still has effects on the current warfare between Ukraine and Russia.

References

- Bohdanova, Tetyana. (2014). Unexpected revolution: The role of social media in Ukraine's Euromaidan Uprising. *European View* 13(1): 133-142.
- Brym, Robert, Godbout, Melissa, Hoffbauer andreas, Menard, Gabe and Tony Huiquan Zhang. (2014). Social media in the 2011 Egyptian uprising. *Media in Society and Sociology* 65(2): 266-292.
- Byun, Chonghyun Christie and Ethan J. Hollander. (2015). Explaining the intensity of the Arab Spring, DOMES. *Digest of Middle East Studies* 24(1): 26-46.
- Carpentier, Nico. (2017). *The discursive-material knot: Cyprus in conflict and community media participation*. New York: Peter Lang.
- Du, Ying Roselyn. (2016). Same events, different stories: Internet censorship in the Arab Spring seen from China. *Journalism and Mass Communication Quarterly* 93(1): 99-117.
- Elena, Martin. (2016). Framing international media in the face of social movements: CNN and Al Jazeera English in the fall of Morsi. *Communication and Society* 29(3): 119-130.
- Heintz, Jim. (2013). Ukraine's Euromaidan: What's in a name? Associated Press, 2 December. Available at: <http://news.yahoo.com/ukraines-euromaidan-whatsname090717845.html>
- Knoblock, Natalia. (2020). *Language of conflict: Discourses of the Ukrainian crisis*. Bloomsbury Publishing.
- Lukin, Annabelle. (2019). *War and its ideologies. A social-semiotic theory and description*. Singapore: Springer, Springer Nature.
- Kapliuk, Kateryna. (2013). Role of social media in EuroMaidan movement essential, *KyivPost*. Available at: <http://www.kyivpost.com/content/ukraine/role-of-social-media-in-euromaidan-movement-essential-332749.htm>
- Lawless, Brandi and Yea-Wen Chen. (2016). Reclaiming their historical agency: A critical analysis of international news discourses on occupy and Arab Spring. *Howard Journal of Communications* 27(3): 185-202.
- Lokot, Tetiana. (2013). As Ukraine's protests escalate, #Euromaidan hashtag in a sea of information. *Global Voices* 6(12). Available at: <http://globalvoicesonline.org/2013/12/06/as-ukraines-protests-escalate-euromaidan-hashtag-lost-in-sea-of-information/>
- Martin, James R. and Peter R. White. (2005). *The language of evaluation: Appraisal in English*. London; New York. Palgrave; Macmillan.
- Onuch, Olga. (2015). Facebook helped me do it: Understanding the EuroMaidan protester 'Tool-Kit'. *Studies in Ethnicity and Nationalism* 15(1): 170-184.
- Partington, Alan. (2015). Corpus-assisted comparative case studies of representations of the Arab World. In *Corpora and Discourse Studies. Integrating Discourse and Corpora*. Paul Baker and Tony McEnery (eds.) London: Palgrave Macmillan, 220-243.
- Said, Edward. (1978). *Orientalism*. New York. Pantheons Books.
- Stubbs, Michael. (2001) Texts, corpora and problems of interpretation: A response to Widdowson. *Applied Linguistics* 22(2): 149-172.

A corpus-assisted discourse analysis of language variation in child-directed praise and reprimands

Freja Verachtert (Katholieke Universiteit Leuven, Belgium)

Eline Zenner (Katholieke Universiteit Leuven, Belgium)

Dorien Van de Mieroop (Katholieke Universiteit Leuven, Belgium)

Keywords: corpus-assisted discourse analysis, language variation, child-directed speech, praise and reprimands, language regards

Through a mixed-methods approach, this study examines language variation in child-directed praise and reprimands with the aim to explore to which extent these specific pragmatic phenomena provide insights into the social meaning of language varieties and caregivers' *language regards* (Preston 2013). In doing so, it builds on work by AUTHORS (2021), who drew a link between the hyperstandardised linguistic situation in the Dutch-speaking community of Belgium and the Western-European ideal of democratic parenting, indicating that caregivers seem to reserve standard language (the "best language" in the speech community, see Van Hoof & Jaspers 2012) for softer control acts (the "best parenting" in a Western-European context, see Pećnik 2007).

Eight Belgian Dutch families with children between 1 and 8 years old were asked to self-record conversations at and around the dinner table in the family home. This resulted in approximately 30 hours of video material, which was transcribed using the Chat conventions of the Childes project (MacWhinney 2000) for the overall transcriptions and the Jeffersonian conventions for the qualitative analyses. The transcriptions culminated in a corpus of 24,903 utterances, comprising a total of 156,415 words of spontaneous speech across the eight families.

To gain a general view on the distribution of standard and non-standard variants in the contexts of praise and reprimands, the corpus is first subjected to a preliminary quantitative variational analysis, focusing on pronouns of address. The outcomes subsequently inform a qualitative discourse analysis of selected cases which are identified through the corpus analyses. This approach enables us to cross-validate our findings and enrich our insights, ultimately providing a more comprehensive understanding of the sociolinguistic implications of variation in child-directed praise and reprimands.

Given the respective positive and negative valence of praise and reprimands and consistent with how caregivers seem to associate the standard variety more with softer control acts than with more direct ones (2021), we expect that caregivers will associate standard forms significantly more with praise than with reprimands. These results may not only provide a new angle in caregivers' language regards as well as on the status of the varieties under scrutiny, they may also underpin the value of incorporating mixed-methods research that specifically combines corpus linguistics and discourse analysis.

References

- MacWhinney, Brian. (2000). *The CHILDES project: The database (2)*. Psychology Press.
- Pećnik, Ninoslava. (2007). Towards a vision of parenting in the best interest of the child. In *Parenting in Contemporary Europe: a Positive Approach*. Mary Daly (ed.) Strasbourg: Council of Europe Publications, 15-36.
- Preston, Dennis R. (2013). The influence of regard on language variation and change. *Journal of Pragmatics* 52: 93-104. 10.1016/j.pragma.2012.12.015.
- Van Hoof, Sarah and Jürgen Jaspers. (2012). Hyperstandaardisering. *Tijdschrift voor Nederlandse taal- en letterkunde* 128(2): 97-125.

Perspectives on democracy and education

A corpus assisted cognitive critical discourse analysis of the Chilean's citizenship education official textbooks

Felipe Villegas (Pontificia Universidad Católica de Chile, Chile)

Keywords: corpus assisted critical discourse analysis, citizenship education, cognitive linguistics, official textbooks

The relationship between democracy and education has been subject to academic discussion for a long time since Dewey's publication of "Democracy and Education" in 1916 (Sant 2019). This educational discussion, which is still relevant and ongoing, has its roots in the different ideological perspectives that struggle to impose their ideas on how democracy should be taught (Apple et al. 2022; Sant 2019).

The literature has grouped the aspects in which these perspectives differ in various continuums (Almeyda & Jiménez 2020; Sant 2019). Among other areas of interest, the literature on Chilean citizenship education has focused on the spectrum between minimalist and maximalist perspectives on democracy and education (Kerr 2002). The government has explicitly mentioned this spectrum in the National Plan for Citizenship Education (Plan Nacional de Formación Ciudadana 2022). However, as the result of the body of research on this area has shown, there is no clarity regarding what perspective is encouraged to be adopted by the educational system. Because of the former, it is possible to observe that different groups of teachers and students may have opposing beliefs and discourses concerning this continuum (Muñoz & Torres 2014; Salinas et al. 2016; Orellana & Muñoz 2019; Soto & Peña 2020; Quilamán & Vitale 2023).

Citizenship education has been a mandatory subject in Chile for the last two high school levels since 2020. With the creation of the subject, the Ministry of Education started the distribution of official textbooks, which are used by 90% of Chilean schools (Olivera 2016); because of the impact these textbooks have on the Chilean educational system, it is relevant to generate research on them.

The present study aims to interpret the discursive construction of Citizenship Education in the official textbooks, specifically in the continuum between minimalist and maximalist perspectives on democracy and education, from a cognitive linguistics approach to critical discourse analysis. To accomplish this objective, it follows Baker's (2023) and Baker and Levon's (2015) methodological recommendations for a Corpus Assisted (Critical) Discourse Analysis combining quantitative techniques in Corpus Linguistics and more qualitative techniques typically associated with discourse analysis.

A small corpus of 294.446 words comprising the four textbooks distributed from 2020 to 2023 was built and stored in SketchEngine. This study operationalises two main differences between the minimalist and maximalist perspectives, namely, the topics related to democracy and the consideration of the student on these topics (Salinas et al. 2016). Wordlists were used to determine the most frequent nouns in the textbooks. The

WordSketch feature was used on these nouns, which may indicate the main topics addressed in the textbooks and then, a Conceptual Frame Analysis (Fillmore 1982, 2006) was carried out based on the revision of concordance lines of these nouns. Regular expressions were used to analyse the proportion of grammatical persons in the sentences in the textbooks (with special attention to the second person), which may contribute, or not, to the conceptualisation of the reader (i.e. the student) under the notion of subjective/objective construal (Langacker 1985, 2007).

The results show that the textbooks give a strong emphasis in teaching rights and construing the students as citizens who are mostly subjects of rights within legal and institutional frames, typically associated with a minimalist perspective. The notion of human rights, associated by the literature with a maximalist perspective (e.g. Almeyda & Jiménez 2020), is, contradictorily, framed within the legal and institutional frames. Consequently, the textbooks discursively construct human rights from a minimalist perspective. Regarding reader construction, only 21 % of the sentences codify the first person plural or the second person, which may promote the conceptualisation of the reader in relation to the topics covered by the textbooks. In conclusion, it is possible to interpret that Chilean citizenship education textbooks may promote a minimalist perspective over a maximalist perspective. However, throughout the textbooks there are also interesting elements from a maximalist perspective evidenced in the analysis.

References

- Almeyda Hidalgo, Liliam and María Soledad Jiménez. (2020). Enseñanza de la historia y concepciones de ciudadanía en las prácticas de profesores en formación. *Sophia Austral* (26): 179-197.
- Apple, Michael W., Gert Biesta, David Bright, Henry A. Giroux, Amanda McKay, Peter McLaren, Stewart Riddle and Anna Yeatman (2022). Reflections on contemporary challenges and possibilities for democracy and education. *Journal of Educational Administration and History*. 54(3): 245-262.
- Apple, Michael W. (2011). Democratic education in neoliberal and neoconservative times. *International Studies in Sociology of Education* 21(1): 21-31.
- Baker, Collin F., Charles J. Fillmore and John B. Lowe. (1998). The Berkeley FrameNet project. En COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics.
- Baker, Paul, Costas Gabrielatos, Majid KhosraviNik, Michał Krzyżanowski, Tony McEnery and Ruth Wodak. (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.
- Baker, Paul. (2006). *Using corpora in discourse analysis*. Continuum
- Baker, Paul. (2023). *Using corpora in discourse analysis*. Bloomsbury Publishing.
- Boas, Hans C. (2021). Construction grammar and frame semantics. In *The Routledge Handbook of Cognitive Linguistics*. Wen, Xu and John R. Taylor (eds.) Routledge, 43-77.
- Evans, Vyvyan. (2019). *Cognitive linguistics: A complete guide*. Edinburgh University Press.
- Fairclough, Norman. (2013). *Critical discourse analysis: The critical study of language*. Routledge.
- Filardo-Llamas, Laura. (2013). 'Committed to the ideals of 1916'. The language of paramilitary groups: the case of the Irish Republican Army. In *Critical Discourse Studies*, 10(1): 1-17.
- Fillmore, Charles J. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*. 280(1): 20-32.

- Fillmore, Charles J. (1982). Frame semantics. In *Linguistics in the Morning Calm*. Hanshin Press, 111-137.
- Fillmore, Charles J. (1985). Frames and the semantics of understanding. In *Quaderni di Semantica*. 6(2): 222-254.
- Fillmore, Charles J. (2006). Frame semantics. *Cognitive linguistics: Basic readings*. 34: 373-400.
- Fillmore, Charles J. (2009). A frames approach to semantic analysis. In *The Oxford Handbook of Linguistic Analysis*. Bernd Heine and Heiko Narrog (eds.) New York. Oxford University Press.
- Hart, Christopher. (2013). Event-construal in press reports of violence in two recent political protests: A cognitive linguistic approach to CDA. *Journal of Language and Politics*. 12(3): 400-423.
- Hart, Christopher. (2014). Discourse, grammar and ideology. *Functional and cognitive perspectives*. Bloomsbury Publishing, 1-256.
- Hart, Christopher. (2020). '28 Palestinians die': a cognitive grammar analysis of mystification in press coverage of state violence on the Gaza border. In *New Directions in Cognitive Grammar and Style*. London. Bloomsbury, 93-116.
- Kövecses, Zoltán. (2020). *Extended conceptual metaphor theory*. Cambridge University Press.
- Langacker, R. W. (1985). Observations and speculations on subjectivity. In *Iconicity in syntax*, 1(1985): 109.
- Langacker, Ronald W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (1). Stanford university press.
- Langacker, Ronald W. (2007). Constructing the meanings of personal pronouns. *Aspects of meaning construction*, 171-187.
- Langacker, Ronald W. (2013). *Essentials of cognitive grammar*. Oxford University Press.
- Marín-Arrese, Juana I. (2011). Epistemic legitimising strategies, commitment and accountability in discourse. *Discourse Studies*. 13(6): 789-797.
- McEnery, Tony and Andrew Wilson. (2001). *Corpus Linguistics: An Introduction*. Edinburgh University Press.
- Oteiza, Teresa. (2018). Prácticas de la memoria en clases de historia: construcción de la evidencia multimodal e intertextual del pasado reciente chileno. In *Discurso & Sociedad* (1): 112-160.
- Oteiza, Teresa. (2023). *What to remember, what to teach- Human Rights Violations in Chile's Recent Past and the Pedagogical Discourse of History*. Equinox.
- Sant, Edda. (2019). Democratic education: A theoretical review (2006–2017). *Review of Educational Research*. 89(5): 655-696.
- Scherman, A., González, R., Browne, M. and Pardo, J. (2022). Encuesta de Opinión Pública, Comparative National Elections Project Chile.
- Unger, Johann W. (2016). The interdisciplinarity of critical discourse studies research. *Palgrave Communications*. 2(1): 1-4.
- Van Dijk, Teun A. (2003). *Racismo y discurso de las élites*. Gedisa.
- Van Dijk, Teun A. (2015). Critical Discourse Analysis. *The Handbook of Discourse Analysis*, 466–485.

Visualising conceptual shift between the religious and the secular in Early Modern English

Chaterine Wong (University of Sheffield, United Kingdom)

Susan Fitzmaurice (University of Sheffield, United Kingdom)

Keywords: historical semantics, Early Modern English, corpus linguistics, data visualisation, linguistic DNA

The Early Modern English period is a time of radical changes. This paper builds on previous studies within the Linguistic DNA project, which applied digital methods in deriving patterns of discursive concepts and modelling semantic change in Early English Books Online (EEBO) (e.g. Fitzmaurice 2022; Fitzmaurice & Mehl 2022; Mehl 2020). The LDNA corpus is built from EEBO which consists of 146,000 titles encompassing a diverse range of discourse genres across English Literature, History, Philosophy, Linguistics, Religion, Politics and Government between 1470-1700.

Close manual analysis of discourse and semantics alone, however, cannot thoroughly discern shifting concepts within the corpus. Distant reading, particularly via lexical co-occurrence data and visualisation, dramatically enriches close reading and manual analysis of discourse and semantics. We experiment with data visualisations; we explore the relationships between a node lemma and its constellation; that is, the sum of the quads (sets of four highly frequent common co-occurring lemmas including the node lemma). We focus on the similarities and differences among the constellations of the node lemmas, *earth*, *heaven*, *church*. To do so, a statistical analysis was conducted in the following steps:

- Calculation of the number of co-occurring unique words, the number of co-occurring quads and their relative frequency around the node lemmas to identify the size and complexity of each constellation and its respective concepts;
- Extraction of shared co-occurring lemmas across these constellations of node lemmas and calculation of their co-occurrence frequency to ascertain their significance in these concept constructions;
- Thematic clustering of lemmas within the constellation of each node lemma to ascertain the composition and distribution of themes which shape the concepts;
- Thematic coding of lemmas within each cluster of the constellations to evaluate the corresponding relationship between concepts and word senses.

The data reveal that in this exemplar, *church* has the highest frequency (1,254,880 out of 1.6 billion tokens; *earth*: 376,877, *heaven*: 682,867). Its constellation is the most complex with the most diverse set of co-occurring unique words (263; *heaven*: 212; *earth*:112). While *church* is composed predominantly of lemmas themed around 'roles and hierarchy' (76.78%), followed by 'nature' and 'name' (8.99% and 8.04%), *heaven* and *earth*

share a similar pattern of composition (*heaven*: 48.91% in ‘nature’, 38.59% in ‘roles and hierarchy’ and 6.74% in ‘supernature’; *earth*: 32.94% in ‘nature’, 31.14% in ‘roles and hierarchy’ and 27.69% in ‘supernature’). The patterns in the example suggest that the concepts around these three lemmas associated with early modern religious concerns are dominated to different degrees by concerns with governance, organisation and structures.

Our results show that by classifying the ideological constructs of these lemmas, we can trace the process of semantic shift of concepts. Data visualisations (e.g. network graphs, sankey diagrams and radial tree diagrams) are used in the process as curation explaining the pattern and as exploration unveiling underlying meanings (Knaflic 2015). This paper demonstrates how data visualisation can be an invaluable tool for representing complex concepts of semantic change.

References

- Fitzmaurice, Susan. (2022). From constellations to discursive concepts; or: The historical pragmatic construction of meaning in Early Modern English. *Transactions of the Philological Society*. 120(3): 489-506.
- Fitzmaurice, Susan and Seth Mehl. (2022). Volatile concepts: Analysing discursive change through underspecification in co-occurrence quads. *International Journal of Corpus Linguistics*. 27(4): 428-450.
- Knaflic, Cole Nussbaumer. (2015). *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons.
- Mehl, Seth. (2020). Discursive quads: New kinds of lexical co-occurrence data with linguistic concept modelling. *Transactions of the Philological Society*. 120(3): 474-488.

Communicating and constructing well-being and human sustainability in corporate discourse

A corpus-assisted approach to CSR communication

Federico Zaupa (Università degli Studi di Modena e Reggio Emilia, Italy)

Keywords: corporate social responsibility (CSR), employees, well-being, corpus-assisted discourse studies (CADS), transparency;

In very recent times, as a result of the mounting pressure on businesses to also achieve goals of social and human sustainability, there exists a growing interest in the field of corporate social responsibility (CSR) to promote employees' well-being. Research carried out from a management perspective has shown that CSR practices oriented towards better income and work-life balance also enhance individuals' work performance and, hence, the quality of the services provided by a company (e.g., Kim & Kim 2020). In other words, studies have suggested a strong link between CSR, perceptions on work meaningfulness, well-being and level of engagement and performance (e.g. Bauer 2022; Cao et al. 2022). However, there is still a paucity of research in corporate discourse focusing on how workers' well-being is linguistically promoted in a transparent way.

Drawing on this background, this paper aims at shedding light on how employees' well-being is discursively constructed and communicated in CSR. A subsequent focus of this research is to assess whether information released by companies in relation to this issue is linguistically transparent or vague (Schnackenberg & Tomlison 2016; Jin 2022). This is done by analyzing a monolingual corpus of 2020-23 CSR disclosures – including CSR/ESG reports, webpages and press releases, all written in English – created by a sample of international companies having their headquarter in different geographical areas. Firms under investigation are chosen among Forbes' "2023 Worlds' Best Employers" and operate in various sectors, such as automotive, banking and financial services, clothing and shoes, drugs and biotechnology, retail and wholesale, software and hardware. The corpus is explored using the tools of corpus-assisted discourse studies (Partington et al. 2013; Marchi & Taylor 2018): these include, more specifically, the identification of the topic-related key words across time, geographical areas and sectors, their collocates and the close reading of their extended concordance lines to observe their lexicophrasological patterns (Sinclair 2003, 2004).

On the one hand, findings suggest that companies tend to highlight the CSR practices for the well-being of their employees, through the prevalent use of promotional lexis and statements of commitment. On the other hand, a closer reading of the extended concordance lines and entire texts reveals a lack of linguistic transparency – especially in terms of information accuracy – with regards to details of actions and outcomes achieved. This is suggested by the co-occurrence of vague lexical items associated with quantity and quality assessment adjectives or verbs. These results might open a new question in relation to the necessity to find more transparent linguistic solutions that companies

should adopt while drafting their documents, according to the readership these texts are address to.

References

- Bauer, Emily Luisa. (2022). Linking perceived corporate social responsibility and employee well-being—A Eudaimonia perspective. *Sustainability*. 14(16): 10240. 10.3390/su141610240
- Cao, Man, Shuming Zhao, Hongjiang Lv and Danxia Wei. (2022). Can we have our cake and eat it too? Corporate social responsibility and employees' eudaimonic well-being and performance. *Business Ethics, the Environment & Responsibility* 31(4): 1040-1052. 10.1111/beer.12449
- Jin, Bixi. (2022). A corpus-assisted study of vague language in corporate responsibility reports of the cosmetics industry. *Ibérica*, 43: 77-102.
- Kim, Minseong and Jihye Kim. (2021). Corporate social responsibility, employee engagement, well-being and the task performance of frontline employees. *Management Decision*, 59(8): 2040-2056. 10.1108/MD-03-2020-0268
- Marchi, Anna and Charlotte Taylor. (2018). *Corpus approaches to discourse: A critical review*. Routledge/Taylor and Francis.
- Partington, Alan, Alison Duguid and Charlotte Taylor. (2013). In *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Vol 55. John Benjamins Publishing.
- Schnackenberg andrew K. and Edward C. Tomlinson. (2016). Organizational transparency: A new perspective on managing trust in organization-stakeholder relationships. *Journal of management* 42(7): 1784-1810. 10.1177/0149206314525202
- Sinclair, John McHardy. (2004). *Trust the text: Language, corpus and discourse*. London. Routledge.
- Sinclair, John McHardy. (2003). *Reading concordances*. London. Longman.

A corpus-based critical discourse analysis on the discursive construction of Chinese national identity in the COVID-19 discourse in Chinese mainstream newspapers and online question-and-answer site

Xiaowen Zhao (University of Birmingham, United Kingdom)

Keywords: corpus linguistics, critical discourse analysis, news media, social media, national identity

National identity refers to national members' emotional attachment and sense of belongingness to a particular national group. It is fluid, unstable, subject to change and can often be aroused by crisis in troubled times (De Cillia et al. 1999; Wodak et al. 2009) such as the global pandemic COVID-19. In this study, I adopt corpus-based critical discourse analysis (CDA) to analyze COVID-19-related news articles and posts on Chinese mainstream newspapers and a Chinese online Q&A forum 'Zhihu' from 27 December 2019 to 7 December 2022. The research aims to uncover how national identities are constructed and negotiated in response to health crises, particularly in fostering a sense of collectiveness, reinforcing group solidarity and promoting health communication while consolidating national unification. By scrutinizing the discursive strategies employed in news media and online forums, we can elucidate the interplay between state power, ideology and the perceptions of ordinary Chinese internet users. The significance of the Q&A forum lies in its role as a platform for ordinary citizens to engage with and challenge dominant discourses. While aligned with mainstream narratives, some posts on Zhihu pose critiques and alternative perspectives, reflecting a nuanced understanding of national identity beyond official rhetoric. Furthermore, exploring the discursive construction of Chinese national identity offers insights into the broader context of international relations, as national identity is intrinsically linked to the Self-versus-Other division.

The analysis is done by examining the collocates and concordances of national identity-related words in the two corpora generated from frequency analysis and keyword analysis (comparing the news corpus against the Zhihu corpus) using Sketch Engine. The concordance analysis is carried out to examine the patterns of use of these collocates in co-text. The analytical frameworks used for CDA are the three-tier framework developed specifically for the analysis on national identity (Wodak et al. 2009) and the Discourse-historical approach.

As for themes related to Chinese national identities, by examining the key terms pertaining to nationhood and their collocates, 'a collective present and future' are expressed through the discourse emphasizing the collective efforts done by Chinese people, reflecting 'the consciousness of the Chinese nation community' advocated by the Chinese government. The themes of 'a homo-nationals' and 'a national body' are expressed in discourse depicting the emotional and behavioral dispositions towards

compatriots in both mainland China, cross-traits and overseas using the strategy of ‘we are all in the same boat’ (Wodak et al. 2009). The theme of ‘a common culture’ are most closely associated with framing on the effectiveness of traditional Chinese medicine (TCM) during covid. Constructive and perpetuation strategies are used to emphasize/presuppose Chinese intra-group sameness, positive political continuation under the leadership of CCP and a positive self-presentation that praises China’s efforts in epidemic prevention and control, in comparison with the other national collectives, especially the negative portrayal of the U.S as China’s enemy (e.g., *the U.S.* tends to collocate with verbs *discrediting* and *accusing* China). The first-person plural pronoun *we* are used to create a ‘we-group’ that encompasses not only Chinese, but also China itself and other nations engaged in cooperative ties with China, serving to build up a group solidarity among Chinese readers while conveying globalist ideology stressing cooperation of mankind. In Zhihu, while most texts carry ideological meanings aligning with those delivered by mainstream newspapers, some pose a challenge to the values and beliefs promoted by the newspapers representing the central of the Chinese government, such as questioning and criticizing the strict or even unreasonable epidemic prevention and control policy.

References

- Rudolf, De Cillia, Reisigl Martin and Wodak Ruth. (1999). The discursive construction of national identities. *Discourse & Society* 10(2): 149–73. 10.1177/0957926599010002002.
- Wodak Ruth, Rudolf De Cillia, Reisigl Martin, Liebhart Karin, Hirsch Aron and Rodger Ruth. (2009). *The discursive construction of national identity* (2). Edinburgh University Press.

The red metaphors in *The Governance of China* and their translation in English

From the perspective of Frame Theory

Yafei Zhu (Shanghai International Studies University, China)

Keywords: red metaphors, Frame Theory, Chinese-English translation, The Governance of China

The Governance of China (GC), a four-volume collection of political articles by China's President Xi Jinping from 2014 to 2022 and translated by the Central Compilation and Translation Bureau in China, serves as the latest authoritative materials for understanding China's spirits. In the context of China's traditional culture, the meaning of red in political texts is not only a color but a symbol of Chinese revolutionary spirit.

Based on the framework of Frame Theory (Fillmore & Atkins 2012), this study presents a cognitively oriented analysis of Chinese red metaphors and their English translation in the Chinese-English parallel corpus of GC, showing how an understanding of cognitive linguistics can be employed by scholars working on translation studies to enhance and support their analysis. Two important corpus tools, AntConc and CUC_ParaConc, are used to retrieve all Chinese words consisting the character “红(red)” in the Chinese corpus and the corresponding English translation of the word in the parallel corpus. In addition, Metaphor Identification Procedure VU University Amsterdam (MIPVU) by Steen et al. (2010) is used to identify different kinds of red metaphors. Followed by that, various translation techniques, methods and strategies (Xiong 2014) are interpreted from the perspective of Frame Theory (Zhou 2008) to reveal the cognitive process in the translation of red metaphors.

The study reveals three key findings. Firstly, the majority of red metaphors in the Chinese volumes appear in the third volume, highlighting 2017-2020 as a pivotal period for promoting the revolutionary spirit in China. However, over half of these metaphors are omitted in the English translation, along with related historical and ideological contexts. Secondly, the translation predominantly adopts a domestication strategy, aiding in the accurate conveyance of Chinese ideologies abroad. In cases of red metaphors that trigger distinct frames, a literal translation approach is often employed for proper nouns, such as *Red Army*, *Red Boat* and *Red Area*. In addition, it is worth noting that the translation of *Hongyan Spirit* into *revolutionary spirit of Hongyan* not only retains the original metaphorical vehicle but also paraphrases its meaning, which effectively transcends the dualistic thinking prevalent in traditional translation studies. Conversely, liberal translation is typically favored for compound terms related to the revolutionary legacy, faith, nature, traditions and resources. Lastly, these differences between the Chinese and English versions are primarily due to the different purposes of promoting the revolutionary spirit within China prior to the CPC's 100th anniversary in 2021 and enhancing international understanding of China.

The study underscores the complexities in translating culturally significant metaphors and highlights the strategic choices made to balance domestic propagation of ideology and international comprehension. It contributes to the field of translation studies by offering a nuanced understanding of metaphor translation in political discourse, particularly in the context of Chinese-English translations.

References

- Lehrer, Adrienne, Eva Feder Kittay and Richard Lehrer. (2012). *Frames, fields and contrasts: New essays in semantic and lexical organization*. Routledge.
- Steen, Gerard., Dorst, A. G., Herrmann, J. B Berenike, Kaal, Anna., Krennmayr, T., & Pasma, T. (2010). A method for linguistic metaphor identification. *Amsterdam: Benjamins*.
- Xiong, Bing 熊兵. (2014). "Fanyi yanjiu zhong de gainian hunxiao: yi 'fanyi celve', 'fanyi fangfa', 'fanyi jiqiao' weili" 翻译研究中的概念混淆——以“翻译策略”、“翻译方法”和“翻译技巧”为例 [Conceptual confusion in translation studies: A case study of translation strategies, methods and techniques]. *Zhongguo fanyi 中国翻译[Chinese Translators Journal]* 35(3): 82–88.
- Zhou, Y 周颖. (2008). "Kuangjia lilun xia de yinyu fanyi" 框架理论下的隐喻翻译 [Metaphor translation in Frame Theory]. *Waiguo yuyan wenxue 外国语言文学[Foreign Language and Literature Studies]* (02): 117120.

Language and gender in food and beverage advertising

Elisabeth Katharina Zimmermann (Universität Innsbruck, Austria)

Anna Islentyeva (Universität Innsbruck, Austria)

Keywords: corpus-assisted linguistic analysis, food and beverage advertising, femininity, (hegemonic) masculinity, multimodal analysis

Advertising is one of the most omnipresent, powerful and persuasive types of discourse. Similar to other types of advertising, food advertising usually employs images of models to promote a company's products, thus creating stereotypically gendered representations.

This study aims to analyse the discursive strategies employed in the representation of gender in a sample of print advertisements launched between 2000 and 2020. This study analyses posters that promote products comprising five major categories: (1) alcoholic and (2) nonalcoholic beverages, (3) fast food, (4) snacks and (5) sport nutrition, in order to trace the relation between the types of endorsed products and the strategies employed in the representations of gender. The brands analysed include Burger King, Carl's Jr., Coca-Cola, Maximuscle, McDonald's, KFC and Nespresso.

The multimodal analysis of advertisements is complemented by a corpus-assisted linguistic analysis of the Corpus of Contemporary American (COCA), which includes an analysis of the selected terms featured frequently in the slogans and taglines of the advertising campaigns in question. For instance, concordance and collocational analyses were conducted on the terms *real man/men*, *strong* and *powerful* as these terms proved to be recurrent in the portrayal of masculinity and are seen as central to its construction.

The key findings show there is a correlation between the representations of masculinity and the notions of strength, (competitive) sports and muscle building (Connell 2005; Connell & Messerschmidt 2005; Katz 2011; Gentry & Harrison 2010; Vokey et al., 2013) in the advertising of sport nutrition (protein-containing and low-fat products) and beverage advertising for both alcoholic (especially beer) and non-alcoholic drinks (Coca-Cola, milk). Meat consumption in fast-food advertising is likewise frequently linked to the idea of a strong man (cf. Adams 2020; Klainberger 202). On the other hand, Coca-Cola intentionally links femininity with Diet Coke and thus with diet and dieting in a more general sense. Finally, the linguistic analysis supports the key findings of the multimodal analysis, which shows that language and advertising operate in a distinct, circular structure and mutually impact one another.

References

- Adams, Carol, J. (2020). *The pornography of meat*. New York: Bloomsbury Academic.
- Connell, Raewyn W. (2005). *Masculinities*. Berkely: University of California Press.
- Connell, Raewyn W. and J. W. Messerschmidt. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society* 19(6): 829–59.
- Gentry, James W. and Robert Harrison. (2010). Is advertising a barrier to male movement toward gender change? *Marketing Theory* 10(1): 74–96.
- Katz, Jackson (2011). Advertising and the construction of violent white masculinity: From BMWs to Bud Light. In *Gender, race and class in media: A critical reader*. Gail Dines and Jean M. Humez (eds.) Thousand Oaks: Sage, 261–9.
- Klainberger, Sarah Agnes Thonsgaard. (2020). The Beef Between Men and Animals: Meat-eating and masculinity in the U.S. – A case study of objectifying fast-food advertising in the Anglosphere. *Leviathan: Interdisciplinary Journal in English*. 6(1) 128–48.
- Vokey, Megan, Tefft, Bruce and Chris Tysiaczny. (2013). An analysis of hyper-masculinity in magazine advertisements. *Sex Roles* 68: 562–576.

German language skills in apprenticeship training

A corpus linguistic discourse analysis of selected scientific journals

Katharina Zipser (Universität Innsbruck, Austria)

Erika Kegyes (Universität Innsbruck, Austria)

Keywords: apprenticeship, language skills, analysis of scientific journals, corpus linguistics, discourse linguistics

What role do language skills play in dual apprenticeship training? – In this article, we address the importance of general, technical and vocational language skills within this context from a quantitative and qualitative perspective.

Various studies deal in particular with the need for German as a second language in the context of vocational training (cf. in particular the early project of the *Deutsche Institut für Erwachsenenbildung* (DIE) *Berufliche Weiterbildung mit MigrantInnen* (1992-1995, cf. Nispel & Szablewski-Çavuş 1996) as well as its later project *Deutsch am Arbeitsplatz* (2007-2011) and follow-up projects) and guidelines have also been published (cf. e.g. *Sprachbedarfsermittlung im berufsbezogenen Unterricht Deutsch als Zweitsprache. Ein Leitfaden für die Praxis* (Weissenberg 2012). Other studies deal with general language needs from a vocational and business education perspective. Efing & Kiefer (2018), for example, provide an overview of the thematization of language and communication in vocational training.

The consideration of general, technical and vocational language skills (cf. for example Kniffka & Roelcke 2016; Roelcke 2020; Roche 2021) is particularly, but not exclusively, important in the context of training apprentices with non-German mother tongue (cf. Dornmayr & Riepl 2023) to prepare them for everyday life and profession. Support in language training during the entire dual training program is deemed important.

In a previous study, we evaluated a large selection of texts from Austrian media (225 different media, evaluated for the period 01/2014-01/2024) to determine to what extent and with which keywords language skills and language teaching are addressed there.

With the same aim in mind, we now examine eight scientific journals, four from the field of vocational training, three from the field of language, DaF/DaZ and one journal that combines the two fields. First, we extract the 100 most frequent lexemes from the individual journals and filter them according to thematic criteria. Then we carry out keyword analyses (cf. Bubenhofer 2013, 2021; Schröter 2019) and determine collocation lists and central multi-word units. Finally, following Spieß 2013, we examine individual texts that were selected with the help of qualitative analyses from a discourse linguistic perspective.

With this approach, we come closer to determining the role language skills play in dual apprenticeship training. Our results shed light on the perspectives, focus and objectives within the scientific discussion.

References

- Bubenkofer, Noah (2021). Exploration semantischer Räume im Corona-Diskurs. In *Sprache in Politik und Gesellschaft. Perspektiven und Zugänge*. Heidrun Kämper and Albrecht Plewni (eds.) Berlin/Boston. de Gruyter, 197-216.
- Bubenkofer, Noah. (2013). Quantitativ informierte qualitative Diskursanalyse. Korpuslinguistische Zugänge zu Einzeltexten und Serien. In *Angewandte Diskurslinguistik. Felder, Probleme, Perspektiven*. Kersten Sven Roth and Carmen Spiegel (eds.) Berlin. Akademie-Verlag. 109-134.
- DIE = Leibniz-Zentrum für Lebenslanges Lernen Bonn im Auftrag des Bundesamtes für Migration und Flüchtlinge, Bonn.
- Dornmayr, Helmut; Riepl, Marlis. (2023). Arbeits- und Fachkräftebedarf/-mangel. In *Österreich 2023*, ibw-summary, Wien.
- Efing, Christian; Karl-Hubert Kiefer (eds.). (2018). In *Sprache und Kommunikation in der beruflichen Aus- und Weiterbildung. Ein interdisziplinäres Handbuch*. Tübingen. Narr.
- Kniffka, Gabriele; Roelcke, Thorsten. (2016). *Fachsprachenvermittlung im Unterricht*, Paderborn. Schöningh.
- Nispel andrea; Szablewski-Çavuş, Petra (1996). Über Hürden, über Brücken. Berufliche Weiterbildung mit Migrantinnen und Migranten. Bonn. Deutsches Institut für Erwachsenenbildung.
- Roche, Jörg. (2021). Berufssprache Deutsch: Bedarf, Konzept und Programmatik. In: *Deutsch als Zweit- und Fremdsprache in Schule und Beruf*. Petra Gretsch and Nadja Wulff (eds.) Boston. Brill, 304-316.
- Roelcke, Thorsten. (2020). *Fachsprachen. 4.*, neu bearb. u. wesentl. erw. Aufl., Berlin. Schmidt.
- Schröter, Melanie. (2019). Vergleichende Analyse von Schlagwörtern in europäischen Migrationsdiskursen. Diskurskontext, Wortgebrauch und Semantik. In: Geranka Rocco and Elmar Schafroth (eds.): *Vergleichende Diskurslinguistik. Methoden und Forschungspraxis*. Berlin. Peter Lang, 359-382.
- Weissenberg, Jens. (2012). *Sprachbedarfsermittlung im berufsbezogenen Unterricht Deutsch als Zweitsprache. Ein Leitfaden für die Praxis*. Hamburg. IQ Fachstelle Berufsbezogenes Deutsch.
- Spieß, Constanze. (2013). Texte, Diskurse und Dispositive. Zur theoretisch-methodischen Modellierung eines Analyserahmens am Beispiel der Kategorie *Schlüsseltext*. In: Roth, Kersten Sven; Spiegel, Carmen (eds.): *Angewandte Diskurslinguistik. Felder, Probleme, Perspektiven*. Berlin. Akademie-Verlag, 17-42.

“Atrocity propaganda” and the language of terror to be found in a German text corpus from 1933.

Hanno Biber (Austrian Academy of Sciences, Austria)

Keywords: digital literary studies, corpus linguistics, critical discourse analysis, historical studies, text corpora

The research question of this proposal is how to use a digital text corpus and how to apply the methods of computational discourse studies and corpus-based text studies for an analysis of the language of terror in Germany in the year 1933. Digital text corpora and digital editions of German original texts of the time will be made use of in order to study the language with a focus on the discourse of threat, terror, violence and related linguistic and discursive strategies of concealing or revealing the intentions of committing atrocities and criminal political aggression. The resources are based upon the text collections of the “AAC Austrian Academy Corpus” and comparable historical text corpora. The research objective is determined by “The Third Walpurgis Night”, the most important contemporary analysis of the language in focus. This historical text, written by the language critic Karl Kraus between May and September 1933, is the most important literary account of the events that happened when Hitler and the Nazis came to power. It explains on three hundred pages the origins, the conditions and the consequences of the unspeakable events that occurred and it gives several hundred of quotations from contemporary sources documenting how the violent actions unfolded and how the crimes were made possible. The text is taken as a starting point for an exemplary corpus-based analysis of the discursive strategies in this particular historical context. The study will also, based upon an index of names, document who was acting, who was suffering or who was witnessing the criminal atrocities committed and how. The source texts, that are registered digitally in the edition, using historical documents and digitalized archive material, will provide detailed examples of the language of terror to be found in the texts from 1933.

References

Biber, Hanno. (2021). „Personenregister“. In: *ACE- Austrian Corpora and Editions: Karl Kraus 1933*, <https://kraus1933.ace.oeaw.ac.at>

Kraus, Karl. (1933). [Anfang Mai 1933], 293 p., Sign. Schwad 01 19 290.1, The National Library of Israel, Jerusalem

The meanings of *crisis*

A corpus-based analysis of parliamentary discourses across space

Ksenija Bogetic (Università degli studi di Ferrara, Italy)

Dario Del Fante (Università degli studi di Ferrara, Italy)

Keywords: crisis, parliamentary discourse, political discourse

The meaning of *crisis* as occurring in today's public discourse has become subject to extensive new theorization (Krzyżanowski et al. 2023; Wodak 2021), not least given the concept's fast shifting connotations amidst the challenges of the pandemic, global conflict and crisis of neoliberalism. Recently, *crisis* has been highlighted as a profoundly discursive and narrative construct that legitimizes re-definitions of social reality (Krzyżanowski et al. 2023), instrumentalized with increasing skill in populist discourses of 'threat', but malleable in various ways in top-down political discourses more broadly.

Departing from this scholarly interest, the present work presents the preliminary results of a wider study aimed at investigating the shifting meanings of *crisis* in different European contexts of political discourse. It applies a combination of corpus-linguistic and Critical Discourse Analysis perspectives (Partington et al. 2013; Baker 2023) to investigate the use of the lemma *crisis*, looking at the discourses of four national parliaments in Europe: UK, Slovenia, Serbia and Italy. As data, we draw on the recently compiled multilingual corpus ParlaMint containing transcriptions of parliamentary (Erjavec et al. 2023). While the study is at this point not primarily comparative, the focus on different socio-political contexts in/beyond the EU and its peripheries is expected to contribute to understanding conceptions of crisis as both globally defined and constrained by local contexts. Specifically, collocation analysis is a productive first step for gaining insights into what is emphasized as *crisis* in the given locales over the past ten years, while also yielding some insights into emerging concepts like *polycrisis* and *permacrisis* that appear to collocate with the node. Qualitative analyses of concordances and broader textual contexts further reveals the various and conflicting interrelations, representations and social associations of *crisis* in the discourses at different points in time. In this preliminary form, the findings will be illustrated through selected corpus-analysis results and discursive examples, while stressing some polarizing representations of 'crisis' most strongly highlighted in our analysis so far.

References

Baker, Paul. 2023. *Using Corpora in Discourse Analysis*. 2nd ed. Bloomsbury Publishing.

Erjavec, Tomaž, Matyáš Kopp, Maciej Ogródniczuk, Petya Osenova, Manex Agirrezabal, Tommaso Agnoloni, José Aires, et al. (2023). *Multilingual comparable corpora of parliamentary debates ParlaMint 4.0*. <https://www.clarin.eu/content/parlamint>, <https://www.clarin.si/repository/xmlui/handle/11356/1859>

Krzyżanowski, Michał, Ruth Wodak, Hannah Bradby, Mattias Gardell, Aristotle Kallis, Natalia Krzyżanowska, Cas Mudde and Jens Rydgren (2023). Discourses and practices of the 'new normal' towards an interdisciplinary research agenda on crisis and the normalization of anti-and postdemocratic action. *Journal of Language and Politics* 22(4).

Partington, Alan, Alison Duguid and Charlotte Taylor. (2013). *Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies (CADS)*. Amsterdam: John Benjamins. 10.1075/scl.55

Wodak, Ruth (2021). Crisis communication and crisis management during COVID-19. *Global Discourse*. 11(3): 329-353.

Code switching between Italian and dialects in TikTok discourse

Claudia Roberta Combei (Università degli Studi di Salerno, Italy)

Luisa Troncone (Università degli Studi di Salerno, Italy)

In the last decades, the Internet has revolutionised communication channels and the dynamics of interaction, with social media standing out as a foremost manifestation of this digital transformation. TikTok, in particular, distinguishes itself from other networking platforms since it prioritises the multimodal dimension of communication, enabled by the instantaneous sharing of user-created videos.

This paper deploys a corpus-assisted approach to investigate the functions of code switching (CS) in the TikTok discourse of Italian users. Here, CS refers to the linguistic phenomenon characterised by the alternating use of Italian language and Italo-Romance dialects by a single speaker within the same speech event, involving the interweaving of passages in both codes (Berruto 1997).

We expect to find various types of code-switching usages on TikTok: defining the linguistic identity of the users; establishing and consolidating social ties in the target linguistic community; and providing a means of accommodation and clarification. Additionally, we hypothesise that CS manifests itself in situations of strong emotionality, in which speakers exert less control over the language. In these instances, slurs and profanity are frequently encountered and when uttered in dialect, they go unnoticed by identification algorithms, avoiding classification as inappropriate content (hate speech, sexually explicit language, etc.).

The ongoing compilation and analysis of the corpus focus on specific elements of TikTok content: the users' oral discourse transcribed with ELAN; automatically generated and manually edited subtitles; and comments associated with the videos. We have extracted, so far, over 20 of such instances of CS that successfully eluded filters designed to identify inappropriate language, such as those in (1), (2) and (3):

(1) ma va Fe kinott [comment to video]

Translation: 'go give some blowjob'.

(2) è semplice: con l'asfaling e il kitammuorting [oral discourse]

Translation: 'it's easy, through bulldozing and cursing their dead relatives'.

(3) accirt [comment to video]

Translation: 'go kill yourself'.

The results will allow us to identify regularities of CS use, considering the context in which this phenomenon occurs. Finally, our study may contribute to research on the creation of language models and algorithms intended to filter inappropriate content on social media.

References

Berruto, Gaetano. (1997). Code switching and code mixing. In *The Dialects of Italy*, edited by Martin Maiden and Mair Parry, 394-400. London and New York: Routledge.

Control procedures in sceptical pandemic discourse

A grounded theory on discursive series and practices

Dario D'Agostino (Zürcher Hochschule für Angewandte Wissenschaften, Switzerland)

Keywords: Foucauldian discourse analysis, grounded theory, corpus linguistics, network analysis

In the context of the COVID-19 pandemic, some scientific disciplines (e. g. epidemiology) have gained particular attention, both in media and in academia. However, the pandemic has also raised numerous questions which can be addressed by linguistic discourse analysis (LDA). In fact, already much research has been done in the field, also regarding discourses in Switzerland¹. To date, however, the *sceptical* pandemic discourse in Switzerland has not been analysed thoroughly from a linguistics perspective.

My doctoral thesis addresses this field of research with analytical (1) and methodological (2) contributions. The dataset consists of an annotated text corpus of 107 sceptical websites, comprising 18.868 texts, 15.108.308 words.

Analytically (1), the reconstruction of discursive practices of the digital sceptical pandemic discourse in German speaking Switzerland stands in the foreground. Particular attention is given to the *discursive control procedures*, with which – according to Foucault² – discourses are organized, controlled, selected and redistributed. By the *principle of commentary*², the discourse is ordered by a gradation among “text”: there are primary, canonical texts which are perpetually recited, repeated and varied by secondary, commentary texts. Current empirical results of my research show a manifestation of the *principle of commentary* in the digital sceptical pandemic discourse: the use of bibliographical referencing and citing, similar to academic conventions and a frequent use of hyperlinks. I claim that these deictic practices have a regulatory function in the discourse. To further investigate on these findings, I have created a deictic network (based on hyperlinks extracted from the project’s corpus) which visualises the digital pointing practices: *who points at whom?* Parallely, patterns of the pointing practices are analysed on the linguistic surface using corpus linguistics methods.

The methodological (2) contribution consists of an adaption of the grounded theory³ approach, which addresses LDA purposes and comprises mapping⁴ and visualization techniques.

¹ cf. e.g. Bubenhofer, “Exploration semantischer Räume im Corona-Diskurs”; Rocco, Dreesen and Krasselt, “Exploration zu Deutschen und italienischen Akteursrollen in Covid-19-Diskursen : Methodologische Überlegungen und praktische Schlussfolgerungen zur vergleichenden Diskurslinguistik.”

² Foucault, *Die Ordnung des Diskurses [Inauguralvorlesung am Collège de France, 2. Dezember 1970]*.

³ Glaser and Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research*.

⁴ Clarke, *Situationsanalyse Grounded Theory nach dem Postmodern Turn*.

References

- Bubenhof, Noah. (2022). Exploration semantischer Räume im Corona-Diskurs In *Jahrbuch des Instituts für deutsche Sprache, Sprache in Politik und Gesellschaft: Perspektiven und Zugänge*. Heidrun Kämper and Albrecht Plewnia. De Gruyter, 197–216.
- Clarke, Adele E. (2012). Situationsanalyse Grounded Theory nach dem Postmodern Turn. In *Interdisziplinäre Diskursforschung*. Juliane Sarnes and Reiner Keller (eds.) Wiesbaden. Springer VS.
- Foucault, Michel. (2017) *Die Ordnung des Diskurses [Inauguralvorlesung am Collège de France, 2. Dezember 1970]*. Frankfurt am Main. Fischer.
- Glaser, Barney and Anselm Strauss. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New Brunswick; London. Aldine Transaction.
- Rocco, Goranka, Philipp Dreesen and Julia Krasselt. (2021). Exploration zu deutschen und italienischen Akteursrollen in Covid-19-Diskursen : Methodologische Überlegungen und praktische Schlussfolgerungen zur vergleichenden Diskurslinguistik. In: *tekst i dykurs - text und diskurs* 15: 287–314. 10.7311/tid.15.2021.11.2021

The making and sharing of news values in online shared political news

The contextual interpretations and influence on news sharing of personalisation and eliteness"

Johanna Disdier (Swansea University, United Kingdom)

Keywords: news values, digital news, corpus linguistics, participatory research, automated content analysis.

Social media have transformed how news is produced and distributed, making users active participants in shaping news narratives. As “co-consumers”, “co-producers” and “producers” (Bruns 2007, Khosravini 2017), they not only share news but also reflect on its values, demonstrating how news values resonate within the digital sphere. From this perspective, the treatment and the selection of news articles are intricately interrelated.

This study focuses on the production and reception of political news in English, shared on Facebook and Twitter, particularly regarding citizen protests following post-pandemic global political elections. The protests, sparked by election outcomes or nominations of former leaders, revealed concerns about potential threats to democracy. The research includes a dataset of 193 shared news articles spanning six countries — United States, Iraq, Peru, Chad, Kenya and Benin — collected between 2021 and 2022 from Buzzsumo. This platform ranks news articles based on social media engagements such as shares, reactions and comments.

This study combines textual analysis, participatory survey methods and automated content analysis to examine how news values Eliteness, referring to stories featuring high-status individuals and Personalisation, referring to stories revolving around the personal experiences of ordinary individuals, are perceived and influence news sharing on social media (Trilling et al. 2016). The relationship between newsworthiness and shareworthiness (Trilling et al. 2016) was explored by considering both user and news characteristics.

First, corpus linguistics tools, specifically frequency and dispersion, were used to identify news values constructed through discourse (Bednarek & Caple 2017). This analysis was supplemented by qualitative methods, including concordance analysis and a survey of 400 participants to delve into perceptions and choices related to news sharing. The study also used automated content analysis of articles and their associated metadata, particularly share counts and news sources.

The results show that there is an overlap of interpretation in constructing and perceiving Personalisation and Eliteness. Personalisation often emerges more in local news, which tends to focus on individual rather than collective experiences. However, when group experiences are mentioned, they are sometimes perceived as personal stories, although the categorisation of Personalisation may be more significant and direct when news narratives are centred on individual experiences rather than those of groups

(Bednarek 2016: 244). In contrast, Eliteness is more prevalent in international news, yet reflects varying political and cultural interpretations of what constitutes an elite figure to a global audience.

Second, Eliteness and Personalisation appear to be predictors of news sharing in several case studies, suggesting that in some contexts, articles that feature elite figures or individuals' personal experiences are more likely to be shared. This influence varies across different regions and audiences, indicating that the impact of these news values on news sharing is context dependent.

References

- Bednarek, Monika. (2016). Investigating evaluation and news values in news items that are shared through social media. *Corpora* 11: 227-257. 10.3366/cor.2016.0093
- Bednarek, Monika and Helen Caple. (2017). *The Discourse of News Values: How News Organizations Create Newsworthiness*. Oxford. Oxford University Press. 10.1093/acprof:oso/9780190653934.001.0001
- Bruns, Axel. (2007). Producers: Towards a Broader Framework for User-Led Content Creation. In *Proceedings of the 6th ACM SIGCHI Conference on Creativity & Cognition*. New York, NY, USA. ACM, 99-106. 10.1145/1254960.1254975
- Khosravinik, Majid. (2017). Social Media Critical Discourse Studies (SM-CDS). In *The Routledge Handbook of Critical Discourse Studies*, edited by John Flowerdew and John E. Richardson. New York. Routledge, 582-596.
- Trilling, Damian, Petro Tolochko and Björn Burscher. (2016). From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics. *Journalism & Mass Communication Quarterly*. 94(1): 38-60. 10.1177/1077699016654682

Silencing climate

A study on televised climate reporting

Irene Elmerot (Stockholms universitet, Sweden)

Andrea Culková (Prešovská univerzita v Prešove, Slovakia)

Irena Reifová (Univerzita Karlova, Czech Republic)

Keywords: climate change, tv, news shows, Czech, corpus-assisted discourse studies

While existing studies extensively assess climate reporting qualitatively, a scarcity persists in quantitative research examining climate change reports on television. Television is still the platform most used worldwide to receive general news (News Source Usage Worldwide 2020, n.d.), climate information included (Painter et al. 2023). Despite the societal importance of this field, television's portrayal of climate change has yet to receive the necessary research attention (Schäfer 2015).

As noted *inter alia* by Schröter and Taylor (2018), the absence and silence of certain topics is still rather under-researched in corpus-assisted discourse studies (CADS), which is one reason why this study brings some important additions to our knowledge of how climate issues are represented in televised media.

Environmental issues are not a rare topic in CADS (e.g. Bednarek et al. 2022), whereas television has not been covered as often. We examine all the news broadcasting on Czech television from January 2012 to April 2022, combining both shorter news reels and longer discussion shows. This corpus is extracted from the archive *Mediasearch*. Drawing on topic modelling followed by keyword and collocation analyses, we scrutinize how an imbalanced prioritization of perspectives collectively impedes the establishment of a nuanced and comprehensive climate change reporting. The corpus contains 532 000 documents. Of these, 0.56 % contain the noun *klima* and 0.62 % contain the adjective *klimatický*, both translated as "climate" in English. The topic models we found in the news do not at all include any established climate change terms, not even "global warming" or "greenhouse effect", which was an unexpected example of silence. As a comparison, Tschötschel et al (2022: 575) showed that the German *Tagesschau* news show mentioned the climate in an average of 20 % of the days 2007–2022. While our analyses thus reveal significant 'climate silence' in the corpus, we also observe a notable silence on social issues and on particular geographical regions. Our poster will reveal some of the issues mentioned instead of climate change and some striking absences.

References

Bednarek, Monika, Andrew S. Ross, Olga Boichak, Y.J. Doran, Georgia Carr, Eduardo G. Altmann, Tristram J. Alexander. (2022). Winning the discursive struggle? The impact of a significant environmental crisis event on dominant climate discourses on Twitter. In *Discourse, Context & Media*. 45: 100564.

News source usage worldwide 2020. (n.d.). Statista. <https://www.statista.com/statistics/198765/main-source-of-international-news-in-selected-countries/> (Retrieved August 27 2023)

Painter, James, Joshua Ettinger, David Holmes, Loredana Loy, Janaina Pinto, Lucy Richardson, Laura Thomas-Walters, Kjell Vowles and Rachel Wetts. (2023). Climate delay discourses present in global mainstream television coverage of the IPCC's 2021 report. *Communications Earth & Environment*. 4(1): 118.

Schäfer, Mike. (2015). Climate change and the media. In *International Encyclopedia of the Social & Behavioral Sciences*. 2(3): 853–859.

Schröter, Melani and Charlotte Taylor (eds.) (2018). Exploring silence and absence in discourse. In *Postdisciplinary Studies in Discourse* 14. Cham: Springer International Publishing.

Tschötschel, Robin, Norman Schumann, Rahel Roloff and Michael Brüggemann. (2022). Der Klimawandel im öffentlich-rechtlichen Fernsehen. *Media Perspektiven*. 12: 574–81.

Unpacking discourses around tobacco product use

A comparative study between the tobacco industry, health bodies and the public

Iona Fitzpatrick (University of Bath, United Kingdom)

Sun Xinmei (University of Bath, UK, United Kingdom)

Keywords: tobacco control, corpus-based approach, corporate discourse, health care, diachronic analysis

Language relating to tobacco products used by transnational corporations, the scientific community and the public do not always overlap (van Druten et al. 2022). Gaps in understanding impact the effectiveness of the communication of health risks to the public and can provide opportunities for the tobacco industry to obfuscate tobacco product-related harms, manipulate public opinions and derail legislative processes relating to tobacco use. (Fitzpatrick et al. 2022, Freeman et al. 2018).

Corpus-based methods have previously been applied to analyses of patient and practitioner language, highlighting how linguistic choices impact dynamics of responsibility in the context of health (Adolphs et al. 2004; Berg et al. 2021). Similarly, differences in the use and understanding of certain phrases relating to public health can stifle efforts to enact effective policy solutions and inhibit behaviour change (amongst tobacco consumers for example).

This project investigates the use of tobacco product-related terms by three impacted groups: the tobacco industry, health bodies and the general public. The primary aim of this work is to contribute to the body of work examining corporate use of language and the role of transnational corporations in the shaping of cultural and behavioural norms.

We use a corpus-based approach to examine two tailor-made corpora and two open-source public corpora representing the three groups: 1) reports and web content produced by the tobacco companies BAT and PMI (~ 6 million words); 2) guidelines produced by the health bodies WHO, CDC, NHS and NICE (~ 2 million words); 3) the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA), representing the general public's language use. We explore differences in how tobacco product-related terms were used between the three groups, how their uses change over time (2003-2023) and assess implications for the negotiation of responsibility and accountability for product use, cessation and regulation. Focusing on terms such as 'smok*', 'quit*', 'cessation', 'switch*', 'vap*', 'addict*' and 'relaps*', we employ techniques including concordances, collocation and usage fluctuation analysis.

The findings revealed disparities between the way tobacco companies and health bodies define and talk about product use. In their narrativisation of cessation-related activities, tobacco companies suggest switching to e-cigarettes is a form of quitting, whereas the health bodies define 'quit' as quitting all forms of tobacco products. Discourses of personal responsibilities to quit/switch and corporate responsibilities to

provide products with a lower risk than cigarettes are found to be prominent in the tobacco industry corpus. Analyses also revealed similarities in the language used by tobacco companies and the general public, especially in relation to vaping. By tracking the fluctuations in frequencies and changes in collocation profiles of the terms of interest, we uncovered key turning points where discourses around newer tobacco products emerge and meaning associations with quitting and product-use begin to change.

Our findings highlight ongoing challenges in the way key groups understand tobacco products and their related behaviours. By uncovering the discrepancies and similarities in how tobacco companies talk about smoking and cessation behaviours compared with health bodies and the public, we have identified areas where improved clarity of messaging, specifically related to shared understandings, is needed.

References

- Adolphs, Svenja, Brian Brown, Ronald Carter, Paul Crawford and Opinder Sahota. (2004). Applying Corpus Linguistics in a health care context. *Journal of Applied Linguistics*. 1(1): 9-28. 10.1558/japl.1.1.9.55871
- Berg, Jonathan, Janneke Harting and Karien Stronks. (2021). Individualisation in public health: reflections from life narratives in a disadvantaged neighbourhood. *Critical Public Health*. 31(1): 101-112. 10.1080/09581596.2019.1680803
- Fitzpatrick, Iona, Sarah Dance, Karin Silver, Marzia Violini and Thomas R. Hird. (2022). Tobacco industry messaging around harm: Narrative framing in PMI and BAT press releases and annual reports 2011 to 2021. *Front Public Health*, 10: 958354. 10.3389/fpubh.2022.958354
- Freeman, Becky, Jo Cranwell, Seydou Drame, Valeska Carvalho Figueiredo, Amr Osman and Anurag Anurag Srivastava. (2018). Expert group report on tobacco advertising, promotion and sponsorship: depiction of tobacco in entertainment media. In *World Health Organisation Framework Convention on Tobacco Control (WHO FCTC)*. <http://www.who.int/fctc/cop/Document-TAPS-1.pdf>
- Vera van Druten, E. A. Bartels, Dike van de Mheen, E. de Vries, A. P. M. Kerckhoffs and Lenny Nahar-van Venrooij. (2022). Concepts of health in different contexts: a scoping review. *BMC Health Services Research*, 22: 389. 10.1186/s12913-022-07702-2

Conceptualizing metaphor use in the Brookings Institution's foreign policy guidance

Katherine Hallin (University of Edinburgh, United Kingdom)

Keywords: metaphor, corpus, cognitive linguistics, foreign policy

Cognitive Linguistics has long viewed metaphor as a means to understanding the make-up and efficacy of political argumentation (Gibbs 2017; Lakoff 2001). However, traditional, mostly qualitative, studies on metaphor in politics lack a comprehensive consideration of metaphors' semantic, grammatical and domain complexity. Results from quantitative and corpus research have only partly verified the conclusions of their predecessors and nonetheless tended to focus on the same small samples of intuited metaphor domains (Charteris-Black 2004, 2011; Feinberg et al. 2020; Musolff 2010, 2021; Thibodeau & Boroditsky 2011). A more recent exception to this trend has been in the data-led identification of metaphors from political corpora. Methods of computational detection and semantic/grammatical classification of metaphor words have documented a wide range of source domain semantics, grammar and connections to target domains (Ahrens & Zeng 2022; Shaikh et al. 2014; Wachowiak et al. 2022).

In this dissertation, I add to this latter literature and answer the question: How might frequent metaphor use in foreign policy guidance entrench concepts in an audience's mind? I visualize metaphor domain complexity in finer detail than previous corpus studies by using a Cognitive Linguistics Word Grammar theoretical framework, based on Gisborne (2010); to illustrate the metaphors I find within a mental network. I account for metaphors' semantic and grammatical diversity with corpus and computational data collection methods to identify a range of real-world source domain words and their context. These are sourced from the foreign policy briefs issued in 2022 by the Brookings Institution think tank, one of the most influential foreign policy think tanks in the world (Barham 2021; McGann 2023).

I assemble a corpus of 319, 571 words by scraping public data from Brookings's website. From this corpus, I find the most frequent metaphor and its semantic and grammatical context using metaphor detection (Wachowiak et al. 2022) and part-of-speech and dependency tagging (Qi et al. 2020). I then draw Word Grammar conceptual networks to visualize the most frequent metaphor, which was found to be 'support' meaning material or political help. These networks illustrate the links that 'support' has with the semantic concepts it inherits from, its proximal grammatical arguments and the metaphorical and literal alternative senses of the 'support' lexeme. This network visualization demonstrates that support's givers and recipients are conceptually closely associated, as is the causal relationship between the sense of 'support', meaning approval and 'support' meaning help. I therefore conclude that repetition of 'support' entrenches its association for Brookings's audience with its political givers, recipients and preconditions.

References

- Ahrens, Kathleen and Zeng, Winnie Huiheng. (2022). Referential and evaluative strategies of conceptual metaphor use in government discourse. *Journal of Pragmatics*. 188: 83–96. 10.1016/j.pragma.
- Barham, James. (2021). *Top influential think tanks ranked for 2023: Academic influence*. <https://academicinfluence.com/inflexion/study-guides/influential-think-tanks>
- Charteris-Black, Jonathan. (2004). *Corpus approaches to critical metaphor analysis*. Palgrave Macmillan, Basingstoke, Hampshire.
- Charteris-Black, Jonathan. (2011) *Politicians and rhetoric: The persuasive power of metaphor* (2). Palgrave Macmillan UK, London.
- Gibbs, Raymond W. (2017). *Conceptual Metaphor Analysis* 17–56. Cambridge University Press. 10.1017/9781107762350.002
- Feinberg, Matthew, Wehling, Elisabeth, Chung, Joanne M., Saslow, Laura R. and Melvæ Paulin, Ingrid. (2020). Measuring moral politics: How strict and nurturant family values explain individual differences in conservatism, liberalism and the political middle. *Journal of Personality and Social Psychology*. 118(4): 777–804 10.1037/pspp0000255
- Gisborne, Nikolas. (2010). *The event structure of perception verbs*. Oxford University Press. *Metaphors of Terror, 2001*. URL <https://press.uchicago.edu/sites/daysafter/911lakoff.html>
- McGann, James G. TTCSP Global go to think tank index reports, 2023. URL https://repository.upenn.edu/exhibits/orgunit/think_tanks
- Musolff andreas. (2010). *Metaphor, nation and the Holocaust: The concept of the body politic*. London: Routledge.
- Musolff andreas. (2021). *National conceptualisations of the body politic : cultural experience and political imagination / Andreas Musolff*. Cultural Linguistics. Springer, Gateway East, Singapore. Qi, Peng, Zhang, Yuhao, Zhang, Yuhui, Bolton, Jason and Manning, Christopher D. Stanza. (2020). A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 101–108. Association for Computational Linguistics. 10.18653/v1/2020.acl-demos.14
- Shaikh, Samira, Strzalkowski, Tomek, Cho, Kit, Liu, Ting, Broadwell, George, Feldman, Laurie, Taylor, Sarah, Yamrom, Boris, Lin, Ching-Sheng, Sa, Ning, Cases, Ignacio, Peshkova, Yuliya and Elliot, Kyle. (2014). *Discovering conceptual metaphors using source domain spaces*. doi: 10.3115/v1/W14-4725
- Thibodeau, Paul H. and Boroditsky, Lera. (2011). Metaphors We Think With: The Role of Metaphor in Reasoning. In *PLoS ONE*, 6(2). 10.1371/journal.pone.0016782
- Wachowiak, Lennart, Gromann, Dagmar and Xu, Chao. (2022). Drum up SUPPORT: Systematic analysis of image-schematic conceptual metaphors. In *Proceedings of the 3rd Workshop on Figurative Language Processing (FLP)*, 44–53, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics. <https://aclanthology.org/2022.flp-1.7>

“[O]f course he’d be confused with all the languages flying about”

A corpus-assisted critical discourse analysis of multilingual parenting discussions on Reddit

Ursula Kania (University of Liverpool, United Kingdom)

Sofia Lampropoulou (University of Liverpool, United Kingdom)

Annarita Magliacane (University of Liverpool, United Kingdom)

Paige Johnson (University of Liverpool, United Kingdom)

Keywords: corpus-assisted critical discourse analysis, social media, multilingualism, raciolinguistics, language ideologies

Reddit is a social media platform with over 100,000 active communities and 57 million daily active unique users (Reddit Press 2021). It provides discussion fora on several topics (‘subreddits’), where registered users (‘redditors’) post content and add upvotes/downvotes to contributions made by others. Although Reddit has received some academic attention in relation to controversial communities (see Chang 2020 on incels), subreddits containing potentially less hostile content are still underexplored.

Taking a corpus-assisted critical discourse approach (Baker & McEnery 2015), the present study involves the analysis of 13 posts and approx. 4000 associated comments on multilingual parenting, taken from the AITA (=‘Am I the Asshole?’) subreddit. AITA is a forum for users to ask questions and discuss moral dilemmas (e.g., ‘AITA for refusing to make my bilingual daughter use an English term for my Brother?’).

The dataset was extracted using `RedditExtractoR` package for R (Rivera 2022). It will be imported to NVivo to carry out basic quantitative analyses (sentiment/word frequencies), followed by in-depth qualitative coding and analysis (for dominant themes/discourses; Fairclough 2010). A particular focus will be on the exploration of the stances users take vis-à-vis language learning and bilingual/multilingual identities, evaluating whether raciolinguistic ideologies are reproduced and/or challenged. A raciolinguistic ideological perspective foregrounds the role that racialised ideological perceptions play in understanding how particular language practices come to be stigmatised, regardless of how well they may correspond to standard forms (Rosa and Flores 2017). For example, white affluent speakers’ bilingualism is often celebrated while that of racialized speakers from low-income backgrounds is stigmatized.

Overall, the results of this study will contribute to and expand on existing research on social media discourse by providing a novel perspective on raciolinguistic ideologies, mono-/multilingualism and language standardisation in online discussion fora.

References

- Baker, Paul and Anthony McEnery (eds.) (2015). *Corpora and discourse: Integrating discourse and corpora*. London: Palgrave.
- Chang, Winnie. (2020). The monstrous-feminine in the incel imagination: Investigating the representation of women as 'femoids' on /r/braincels. *Feminist Media Studies*. 22(2): 254-270. 10.1080/14680777.2020.1804976
- Fairclough, Norman. (2010). *Critical Discourse Analysis: The critical study of language*. 2nd ed. New York: Longman.
- Rivera, I. (2022). 'Package 'RedditExtractoR'. Available at: [RedditExtractoR: Reddit Data Extraction Toolkit \(r-project.org\)](https://github.com/RedditExtractoR/Reddit-Data-Extraction-Toolkit).
- Rosa, Jonathan and Nelson Flores. (2017). Unsettling race and language: Toward a raciolinguistic perspective. *Language in Society* 46(5): 621-647.

Social knowledge in controversial discourses

Challenges in annotating and automating topoi

Carina Kiemes (Technische Universität Darmstadt, Germany)

Marcus Müller (Technische Universität Darmstadt, Germany)

Keywords: discourse analysis, collaborative annotation, classification, automation, large language models

In the proposed paper, we will present and discuss the results of a project on the annotation and automation of interpretative topoi. We define interpretative categories as those that do not necessarily appear on the linguistic surface, as implicit aspects as well as the contextual knowledge of the annotator are important for the annotation (Bender et al. 2023).

The project is part of a research group that analyses controversial political discourses in Germany since 1990. Its methodological program is to improve corpus linguistic methods of discourse analysis with the help of expert knowledge from qualitative research. In the process, a tag set consisting of collaboratively developed categories is being created, from which researchers choose relevant categories for their analysis and annotate their data. Semantic and pragmatic categories are annotated manually. Automation experiments are carried out with selected sentences of annotated data in order to be able to apply the corresponding tags to the entire corpus.

However, tag set creation, collaborative annotation and automated classification pose challenges. The creation of a tag set and the collaborative annotation of the project data are characterised by a constant process of negotiation and the externalisation of knowledge. Those categories that are used in several subprojects must be negotiated by the annotators to find commonalities across sub-discourses, despite the differences in thematic orientation. Furthermore, the annotators must agree on certain criteria to ensure the comparability of the annotations, such as segmentation standards and depth of interpretation. The comparability of the annotations and the distinctiveness of the categories are essential, especially for the automated classification of the data.

As a practical use case, we will discuss the category of topoi. Topoi, as we use the term, refers to the essential function of creating 'acceptable content relations between argument and conclusion' (Kienpointner 1992: 43), a process integral to establishing the credibility and acceptability of an argument's premise and its consequent thesis. Relationships that are viewed as acceptable can provide information about shared social knowledge. Using this particular example, we inform about our findings in collaborative negotiation processes in the annotation of topoi, e.g., segmentation standards and the depth of interpretation. Furthermore, we show the results of the automated classification of the topos of autonomy using a data set from the subproject on *Human and Technology* with a thematic focus on bioethics. Our data consists of linguistically preprocessed corpora from three main sources:

- the plenary minutes of the German Bundestag (Müller 2022),
- German news media, e.g., FAZ, Die Zeit, taz, BILD,
- information brochures or position papers on bioethics.

To classify the topoi of autonomy in the data, we work with the pretrained transformer language model BERT (Devlin et al. 2019) and use our linguistic background to evaluate the results. Based on current experiments, we expect positive results in the automated classification of topoi, especially through the additional annotation of phrases that the annotators consider to be crucial for their annotation decision, i.e., a human “attention mechanism”.

References

- Bender, Michael, Maria Becker, Carina Kiemes and Marcus Müller. (2023). Category Development at the Interface of Interpretive Pragmalinguistic Annotation and Machine Learning – Annotation, Detection and Classification of linguistic routines of discourse referencing in political debates. *Digital Humanities Quarterly Special Issue Working on and with Categories for Text Analysis: Challenges and Findings from and for Digital Humanities Practices*. 17(3) <http://www.digitalhumanities.org/dhq/vol/17/3/000720/000720.html>
- Devlin, Jacob, Ming-Wei Chang, Kenton Lee and Kristina Toutanova. (2019). BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. arXiv. <http://arxiv.org/abs/1810.04805>
- Kienpointner, Manfred. (1992). *Alltagslogik. Struktur und Funktion von Argumentationsmustern*. Stuttgart-Bad Canstatt: Frommann-Holzboog.
- Müller, Marcus. (2022). Die Plenarprotokolle des Deutschen Bundestags auf Discourse Lab. *Korpora Deutsch als Fremdsprache*, 2(1): 123–127. 10.48694/kordaf-3492

Densification in scientific texts

Exploring phrasal complexity

Isabell Landwehr (Universität des Saarlandes, Germany)

Keywords: historical language change, corpus linguistics, register, written scientific English discourse, complexity

We analyze the use of nominal compounds (i.e. nouns modified by one or several nouns) in a historical corpus of written scientific English discourse. Nominal compounds are a key feature of scientific English writing (e.g. Halliday 1988) and a form of phrasal elaboration (e.g. Biber & Gray 2010). They are highly compressed structures, allowing much information to be transmitted at once. With scientific writing becoming increasingly compressed in the noun phrase, the use of nominal compounds has increased significantly over the last two centuries (e.g. Hundt et al. 2012). However, accounts like Uniform Information Density theory (Jaeger 2010) posit that language users aim to distribute information uniformly along the linguistic signal. To investigate if this principle applies to scientific English discourse, we analyze if the use of these compressed structures is balanced out by the use of less condensed structures, such as finite complement clauses.

We analyze how different factors influence phrasal elaboration, using mixed effects regression models implemented in R (R Core Team 2023). Our dataset is the Royal Society Corpus (Fischer et al. 2020; Menzel et al. 2021), a diachronic corpus containing articles, reports and other publications of the Royal Society. We focus on texts from the second half of the 20th century, examining how structural compression affects the number of dependent clauses. We also consider the effects of sentence length, time and text type on densification and use the information-theoretic notion of surprisal (Shannon 1948), which is the predictability of a linguistic unit in a certain context (here: the predictability of a word based on the three preceding words).

Our preliminary findings suggest that compounds are associated with fewer dependent clauses in the same sentence. We also find that the likelihood of compounds increases with greater sentence length and over time, while it decreases with some text types. In the next step, we explore the interaction of phrasal elaboration with predictability. We expect compounds to be preferably used in more predictable contexts.

References

- Douglas Biber and Bethany Gray. 2010. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes*. 9: 2-20.
- Stefan Fischer, Jörg Knappen, Katrin Menzel and Elke Teich. (2020). The Royal Society Corpus 6.0: Providing 300+ Years of Scientific Writing for Humanistic Study. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Marseille, France. European Language Resources Association, 794-802.
- M.A.K. Halliday. (1988). On the language of physical science. In *Registers of Written English: Situational Factors and Linguistic Features* edited by Mohsen Ghadessy, 162-178. London/New York. Pinter Publishers.
- Marianne Hundt, David Denison and Gerold Schneider. (2012). Relative complexity in scientific discourse. *English Language and Linguistics* 16(2): 209-240.
- T. Florian Jaeger. (2010). Redundancy and reduction: Speakers manage syntactic information density. In *Cognitive Psychology* 61: 23-62.
- Katrin Menzel, Jörg Knappen and Elke Teich. (2021). Generating Linguistically Relevant Metadata for the Royal Society Corpus. *Research in Corpus Linguistics* 9(1): 1-18. 10.32714/ricl.09.01.02
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- C.E. Shannon. (1948). A mathematical theory of communication. *Bell System Technical Journal*. 27: 379-423.

Disciplines, demographics and attainment

Creating corpora of student writing and lecturer feedback

Maria Leedham (The Open University, United Kingdom)

Dana Therova (The Open University, United Kingdom)

Jackie Tuck (The Open University, United Kingdom)

Thomas Ullmann (The Open University, United Kingdom)

Prithvi Shrestha (The Open University, United Kingdom)

Keywords: student writing, lecturer feedback, interviews, disciplinary differences, demographic differences

Distance learning in Higher Education has become increasingly prevalent, particularly post-Covid 19 (Dos Santos 2022), yet little is known about the linguistic features of distance students' assessed writing or about lecturer feedback on their texts. Much recent work is based solely on students' writing and does not consider feedback comments (e.g. Nesi & Gardner 2012). The current project explores both student writing and lecturer feedback in a distance learning setting across four disciplines (Business, Engineering, History and Childhood Studies) with a focus on how linguistic features vary across disciplines, demographic student profiles and levels of attainment. Also explored is how lecturers respond to student writing, with the aim of pointing to ways of reducing student attainment gaps.

So far we have compiled an innovative 4-million-word corpus of over 2,500 undergraduate assignments written in a distance learning context over a five-year period (2017-2022) from a total of 192 students across four disciplines. Student demographic data such as disability, carer status and socio-economic status has also been collected in order to compare texts across categories. In addition, a 0.5-million-word corpus of lecturer feedback from on-script comments has been compiled. In the next phase of the project (January-June 2024), corpus analysis of both student texts corpus and feedback corpora will be complemented by 'talk-around-text' interviews (Lillis 2009) exploring lecturers' practices and perspectives.

By the time of the conference we will be able to display results around a wide range of linguistic features across a variety of subcorpora and potentially also to track the potential impact of extensive/less extensive feedback comments. We will have completed interviews with tutors and gained insight into how assessed writing is read and evaluated, potentially identifying links between particular textual features and positive or negative evaluations.

References

Dos Santos, Luis M. (2022). Online learning after the COVID-19 pandemic: Learners' motivations. *Frontiers Education*. 7: 1-13.

Lillis, Theresa. (2009). Bringing writers' voices to writing research: Talk around texts. In *Why writing matters: Issues of access and identity in writing research and pedagogy*. *Studies in written language and literacy* (12). Awena Carter, Lillis Theresa and Sue Parkin (eds.) Amsterdam: Benjamins, 169–187.

Nesi, Hilary and Sheena Gardner. (2012). *Genres across the disciplines: Student writing in higher education*. Cambridge: Cambridge University Press.

Echoes of manipulation

Corpus insights into narcissistic abuse discourses

Morana Lukač (Rijksuniversiteit Groningen, Netherlands)

Keywords: narcissistic abuse, trauma recovery, collocational analysis, podcast analysis

Narcissistic abuse, the type of abuse inflicted by individuals exhibiting severe narcissistic traits, which is linked to intimate partner violence, domestic abuse and coercive control (Howard 2022: 84), affects an estimated one in five individuals (Brown 2010). Despite its prevalence, the omission of narcissistic abuse from current psychological diagnostic classifications, including the latest edition of the DSM-V (American Psychiatric Association 2022), has arguably impeded comprehensive research in this domain.

This study employs a corpus-assisted discourse analytical approach to explore 105 narratives from the *Narcissist Apocalypse* podcast (2019–2023), shedding light on the underexplored realm of recovery from narcissistic abuse. Examining the recovery group's insider vocabulary, specifically the manipulative strategy of love bombing, reveals its pervasive presence in 90% of romantic relationships with narcissists.

Collocational analysis uncovers interconnected manipulative strategies such as mirroring and future faking, providing a more nuanced understanding of assumed narcissists' tactics. The narratives illustrate a two-stage relationship pattern: an initial trust-building phase involving love bombing, mirroring and future faking, followed by gaslighting and devaluation.

Gender disparities identified in support group discussions underscore the need for further research to unveil potential gendered differences in abuse survivors' language use. The study also hints at variations in strategies employed by narcissistic parents compared to narcissistic romantic partners, urging focused exploration into childhood victims of narcissistic abuse.

This study bridges lived experiences with abuse research, identifying patterns in abusive relationships. The insights not only guide future research directions but also foster a nuanced understanding of narcissistic abuse dynamics within the examined corpus, addressing a crucial research gap in this critical area of mental health.

References

- American Psychiatric Association (2022). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association.
- Brown, Sandra L. (2010). 60 million people in the U.S. negatively affected by someone else's pathology. *Psychology Today* <https://www.psychologytoday.com/us/blog/pathological-relationships/201008/60-million-people-in-the-us-negatively-affected-someone-elses> (Accessed 12 January 2024).
- Howard, Vickie. (2022). (Gas)lighting their way to coercion and violation in narcissistic abuse: An autoethnographic exploration. *Journal of Autoethnography*. 3(1): 84–102. 10.1525/joae.2022.3.1.84

“Taking chronic recidivists off the street would protect lives”

Corpus linguistic ventures into criminology

Karoline Marko (Universität Graz, Austria)

Keywords: forensic linguistics, criminology, New York City, news discourse

Mass incarceration and high rates of recidivism in the US (e.g. Ostermann & Hasimi 2022) have long been discussed in literature. While the prevalent punitive culture in the US has fostered ever harsher punishments (e.g. Darley 2005), research evidence indicates that in fact lighter punishments have positive effects on recidivism rates (e.g. van Ginneken & Palmen 2023). One endorsing aspect of recidivism is incarceration itself, as prisons are considered criminogenic environments (e.g. Harding et al. 2017). Also, even after release, the label of ‘criminal’ remains. In criminology, one theory that overlaps with linguistics is that of Labelling Theory (Tannenbaum 1938), which argues that the label a person is given (e.g., “criminal”, “inmate”) determines the self-identity of that person, as he or she will incorporate the qualities of the given label and act accordingly (Singelstein & Kunz 2021).

To investigate the phenomenon of recidivism in the news, a corpus of 666,290 words comprised of articles of the *New York Times* (NYT) and the *New York Post* (NYP) was used. The study aims to highlight how the label “recidivist” is used in the reporting of the respective newspapers, how this is connected to underlying (political) ideologies and what the real-life consequences of such reporting and labelling can be. These aims are achieved by investigating word lists, keyword lists, collocations and concordances.

The analysis shows that conservative political ideologies tend to give rise to stronger punitive attitudes that connect recidivism to a person’s character rather than regarding it as a problem of the criminal justice system. This type of reporting is found more in the NYP, while the reporting of the NYT is centered around measures that can reduce recidivism. The NYP also occupies itself with terrorist recidivism, while the NYT exhibits a more differentiated style of reporting.

References

- Darley, J. (2005). On the unlikely prospect of reducing crime rates by increasing the severity of prison sentences. *Journal of Law and Policy*. 13(1): 189-208.
- Harding, David J., Jeffrey D. Morenoff, Anh P. Nguyen and Shawn D. Bushway. (2017). Short- and long-term effects of imprisonment on future felony conviction and prison admissions. *PAS* 114(42): 11103-11108.
- Ostermann, Michael and Sadaf Hashimi. (2022). Recidivism among people convicted of gun offenses: a call to better leverage reentry resources to decrease gun violence. *Justice Quarterly*. 40(6): 1-22.
- Singelstein, Tobias and Karl-Ludwig Kunz. (2021). *Kriminologie. Eine Grundlegung*. Bern: Haupt.
- Tannenbaum, Frank. (1938). *Crime and the community*. Boston. MA. Ginn.
- van Ginneken, Esther FJC and Hanneke Palmen. (2023). Is there a relationship between prison conditions and recidivism? *Justice Quarterly*. 40(1): 106-128.

Promises and pitfalls

Dissecting open science discourses

Ulrike Pfeifer (Technische Universität Dresden, Germany)

Keywords: digitization and societal change, open science, discourse analysis, mixed methods

The ongoing digitization and emergence of Web 2.0 have sparked a radical shift in scientific production and communication, known as Open Science. This movement, which is symptomatic of the “Politics of Openness” (Hartley et al. 2018), responds to a perceived crises in science (cf. Mirowski 2018: 174–187) and aims for a more trustworthy, efficient and democratic scientific process. However, my research questions whether the discourse surrounding Open Science inadvertently adopts economic and neoliberal ideologies, potentially undermining its intended goals.

The study is based on a corpus, that covers a period of 16 years and consists of:

1) Participation-oriented discourse realizations (Roth 2015), that are elicited in the form of semi-structured interviews conducted with members of a grassroots Open Science Initiative at TUD Dresden University of Technology and staff members working in Open Science related departments of the Sächsische Landesbibliothek – Staats- und Universitätsbibliothek Dresden (SLUB).

2) Publications of said initiative and library (mission- and transparency statements, strategy papers, blog entries, periodical articles, third-party funding applications and annual reports).

The analysis primarily attends to the lexical and argumentative level of discourse, combining quantitative and qualitative methods (Bubenhofer 2013). Using the SketchEngine (Kilgarrieff et al. 2004) keywords are being calculated and their persuasive potential is being analyzed. Additionally, in a more qualitative approach, salient metaphors and their argumentative function are being examined.

For example, the domain of movement metaphors proves to be very productive throughout the entire corpus, e.g. when open, replicable science is described as a *motor* for the *progress of knowledge* (“Erkenntnisfortschritt”) or the *faster growth of knowledge*, thus functioning as an argument in favor of Open Science. This corresponds with the keyword analysis, which reveals several lexemes that refer to concepts of movement and growth (e.g. *expansion, goal, growing, transformation, progression, dynamic*).

Paradoxically, while interviewees mentioned the increasing pressure to perform and produce as an obstacle to trustworthy science, capitalist logics of growth and acceleration are invoked to argue for the application of Open Science practices, that are meant to cure the ailments of the science system.

The presentation draws on research conducted in the project “Digitalisierung als Disruption von Wissenssystemen – Open(ing) Knowledge (DiaDisk)” that is part of the EXU measure “Disruption and Societal Change (TUDiSC)” at TUD Dresden University of Technology.

References

- Bubenhof, Noah. (2013). Quantitativ informierte qualitative Diskursanalyse: Korpuslinguistische Zugänge zu Einzeltexten und Serien. In *Angewandte Diskurslinguistik*, Kersten Sven Roth and Carmen Spiegel (eds.) Akademie Verlag, 109–134. 10.1524/9783050061054.109
- Hartley, Sarah, Sujatha Raman, Alexander Smith and Brigitte Nerlich. (2018). *Science and the politics of openness: Here be monsters*. Manchester University Press.
- Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž and David Tugwell. (2004). The sketch engine. In *Proceedings of the 11th EURALEX International Congress*, 105–116.
- Mirowski, Philip. (2018). The Future(s) of open science. In *Social studies of science*. 48(2): 171–203. 10.1177/0306312718772086
- Roth, Kersten Sven. 2015. Diskursrealisationen: Grundlegung und Methodischer Umriss Einer Pragmatisch-Interaktionalen Diskurssemantik. *Philologische Studien und Quellen Band 247*. Berlin: Erich Schmidt Verlag.

Lexicographer as a stylist

Challenges and possibilities to include LLMs and corpus query tools in lexicographic work

Lydia Risberg (Institute of the Estonian Language, Estonia)

Maria Tuulik (Institute of the Estonian Language, Estonia)

Kristina Koppel (Institute of the Estonian Language, Estonia)

Esta Prangel (Institute of the Estonian Language, Estonia)

Eleri Aedmaa (Institute of the Estonian Language, Estonia)

Margit Langemets (Institute of the Estonian Language, Estonia)

Sirli Zupping (Institute of the Estonian Language, Estonia)

Ene Vainik (Institute of the Estonian Language, Estonia)

Sven-Erik Soosaar (Institute of the Estonian Language, Estonia)

Hanna Pook (Institute of the Estonian Language, Estonia)

Keywords: lexicography, Large Language Models, corpus query tools, registers

Could a lexicographer be a stylist? If yes, which tools do they need? In the poster presentation, we outline the initial findings regarding the usage of Large Language Models (LLMs) and corpus query tools for register identification. To accurately determine the register of specific words it is crucial to understand the context in which these words are typically used (Biber & Conrad 2009). Our initial objective is to analyze the actual usage of words previously categorized as colloquial or informal by Estonian language planning. We aim to test the possibility of automatically detecting registers with LLMs and corpus query tools to add labels to a dictionary based on actual language usage. To identify registers, the Estonian National Corpus (in Sketch Engine, Kilgarriff et al. 2014) employs a hierarchical classification of texts (Koppel & Kallas 2022). We compare the effectiveness of LLMs with the outcomes of corpus query tools and the expertise of L1 lexicographers.

In Estonia, the general understanding of language is strongly influenced by Standard Language Ideology, i.e. the belief that one language variant is better than others (Lindström et al. 2023). This research is motivated by the Estonian discourse where a heated discussion has been going on (see Koreinik 2023), including the need to determine register labels in a dictionary. Even L1 language professionals want more confidence whether a word or its meaning is appropriate for Standard Estonian. The user experience survey on the language portal *Sõnaveeb* ('Word Web', Tavast et al. 2018; Koppel et al. 2019), conducted by the Institute of the Estonian Language (EKI) in 2023, revealed the same attitude (Langemets et al. 2024).

The corpus and corpus query tools have been intensively used for compiling the *EKI Combined Dictionary* (since 2019, displayed via *Sõnaveeb*, see Langemets et al. 2021), but LLMs have not been tested so far (unlike elsewhere: TLex, see Joffe et al. 2003). The

research project focuses on potential applications of LLMs in lexicographical work to find new ways to make lexicographer's work more efficient.

References

- Biber, Douglas & Susan Conrad. (2009). *Register, genre and style*. Cambridge. University Press. 10.1017/CBO9780511814358
- Joffe, David, Gilles-Maurice de Schryver & Daniel J. Prinsloo. (2003). Computational Features of the Dictionary Application TshwaneLex. *Southern African Linguistics and Applied Language Studies*. 21(4): 239–250.
- Kilgarriff, Adam, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý & Vít Suchomel. (2014). The Sketch Engine: Ten years on. *Lexicography*: 1: 7–36. 10.1007/s40607-014-0009-9
- Koppel, Kristina, Arvi Tavast, Margit Langemets, Jelena Kallas. (2019). Aggregating dictionaries into the language portal Sõnaveeb: issues with and without a solution. In: *Proceedings of the eLex 2019 conference* Kosem, I., Zingano Kuhn, T., Correia, M., Ferreria, J. P., Jansen, M., Pereira, I., Kallas, J., Jakubíček, M., Krek, S. & Tiberius, C. (eds.). Sintra, Portugal. Brno: Lexical Computing CZ, 434–452.
- Koppel, Kristina & Jelena Kallas. (2022). Eesti keele ühendkorpuste sari 2013–2021: mahukaim eesti keele kogu. Margit Langemets, Maria-Maren Linkgreim & Helle Metslang (toim). *Eesti Rakenduslingvistika Ühingu aastaraamat*. 18: 207–228. 10.5128/ERYa18.12
- Koreinik, Kadri. (2023). Extra-linguistic arguments in the XXI century corpus and status planning: Superdictionary in times of language standardization from below. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-Ugric Linguistics* 14(1): 117–148. 10.12697/jeful.2023.14.1.04
- Langemets, Margit, Kristina Koppel, Jelena Kallas & Arvi Tavast. (2021). Sõnastikukogust keele portaaliks. *Keel ja Kirjandus*. 8–9: 755–770. 10.54013/kk764a6
- Langemets, Margit, Lydia Risberg, Tiina Paet & Kristel Algvere. (2024). *Sõnaveebi kasutajauuring 2023. Raport*. Eesti Keele Instituut. https://portaal.eki.ee/images/EKI_dokumendid/sonaveebi%20kasutajauuring_2023.pdf
- Lindström, Liina, Lydia Risberg & Helen Plado. (2023). Language ideologies and beliefs about language in Estonia and Estonian language planning. *Eesti ja Soome-Ugri Keeleteaduse Ajakiri. Journal of Estonian and Finno-ugric Linguistics*. 14(1): 7–48. 10.12697/jeful.2023.14.1.01
- Tavast, Arvi, Margit Langemets, Jelena Kallas & Kristina Koppel. (2018). Unified Data Modelling for Presenting Lexical Data: The Case of EKILEX. *Proceedings of the XVIII EURALEX International Congress: EURALEX: Lexicography in Global Contexts*. Jaka Čibej, Vojko Gorjanc, Iztok Kosem and Simon Krek (eds.) Ljubljana: Ljubljana University Press, 17-21.

Anonymous discussion and text visualizing

A case study of YouTube comments on the Ukrainian war

Akira Sano (University of Tsukuba, Japan)

Keywords: text mining, Russian study

This study analyses Russian-language comments posted on YouTube videos about the war in Ukraine by using text mining methods. The analysis results suggest the possibility to influence what individuals think and believe during the war by changing the context in digital public space.

This study visualizes the YouTube comments about Ukrainian war with text mining to figure out what Russian people react to this war. Combining text analytics and political communication studies can reduce ideological biases in discussions on international conflicts. In addition, anonymous comments provided by YouTube are well suited to sensitive topics such as war.

The analysis subject is Russian comments posted on the 10 most-viewed YouTube videos collected by searching videos for “украина (Ukraine)” and “война (war)” and filtering the results by highest view count. This study conducts two types of text mining methods such as co-occurrence analysis and LSS. Co-occurrence analysis visualises the content of a sentence by connecting words that occur simultaneously in a sentence with a line. LSS is a type of semi-supervised machine learning model employing word-embedding techniques and classify words in the corpus to “negative” or “positive” categories.

Results (Co-occurrence Analysis):

In Figure 1, high frequency words such as 'Russian', 'Ukraine' or 'War' are plotted at the center of the terms network. At the left side “Human” and “Life” related words are displayed such as 'human', 'family', 'Christ' or 'time'. At the right side “Country and Place Name” related words are displayed such as 'USA', 'Kiev', 'West' or 'Poland'.

Results (LDA analysis):

Figure 2 shows “Military and War” related words are given low polarity scores and classified as negative such as 'strike/attack', 'submarine' and 'ballistic missile'. In comments regarding China, Figure 3 shows “Economy and bilateral” related words are given high polarity scores and classified as positive such as 'corporation', 'locomotive' and 'intimate'.

It turns out Russian users tend to associate themselves with "connections between people" or "faith in God" when referring to the war in Ukraine, but the users display a more aggressive attitude toward USA by mentioning specific names of strategic missiles or military weapons in comments regarding USA. When it comes to China, Russian users are likely to hope to cooperate on economic and business areas with China instead of

military assistance. The analysis results suggest the possibility to figure out what individuals in Russia feel about USA and expect from China in the Ukrainian war.

References

Benoit, Kenneth, Kohei Watanabe, Haiyan Wang, Paul Nulty, Adam Obeng, Stefan Müller and Akitaka Matsuo. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*. 3(30): 774.

Watanabe, Kohei, Elad Segev and Atsushi Tago. (2022). Discursive diversion: Manipulation of nuclear threats by the conservative leaders in Japan and Israel. *International Communication Gazette*. 7-8(84): 721-748.

Appendix

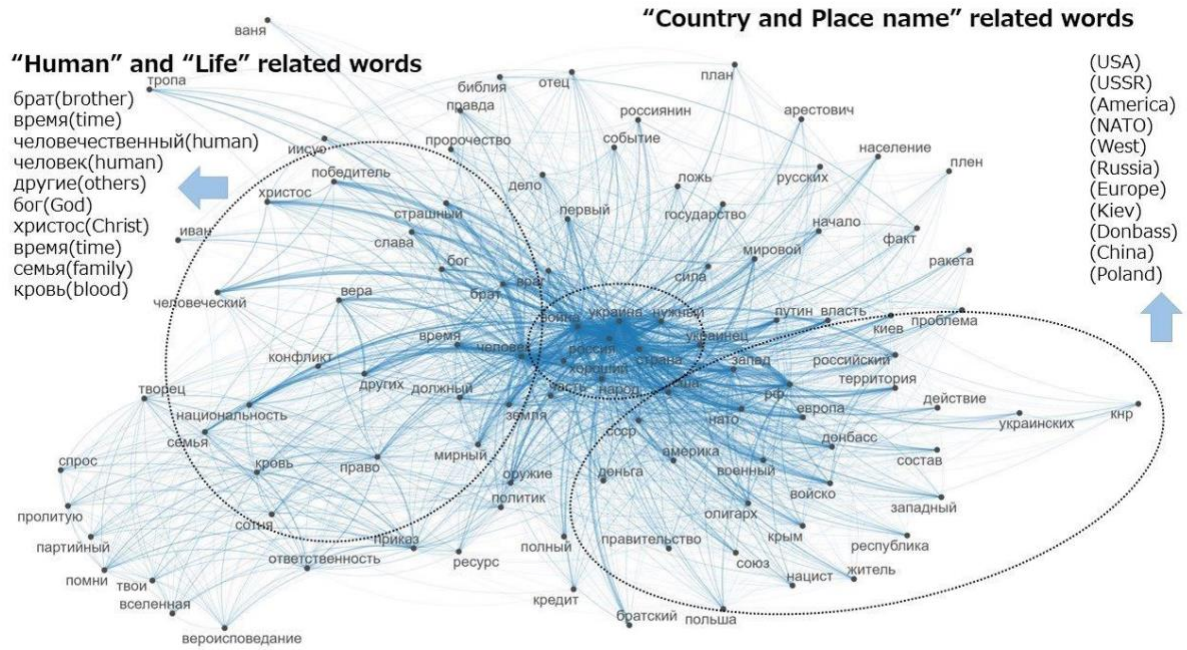


Figure 1

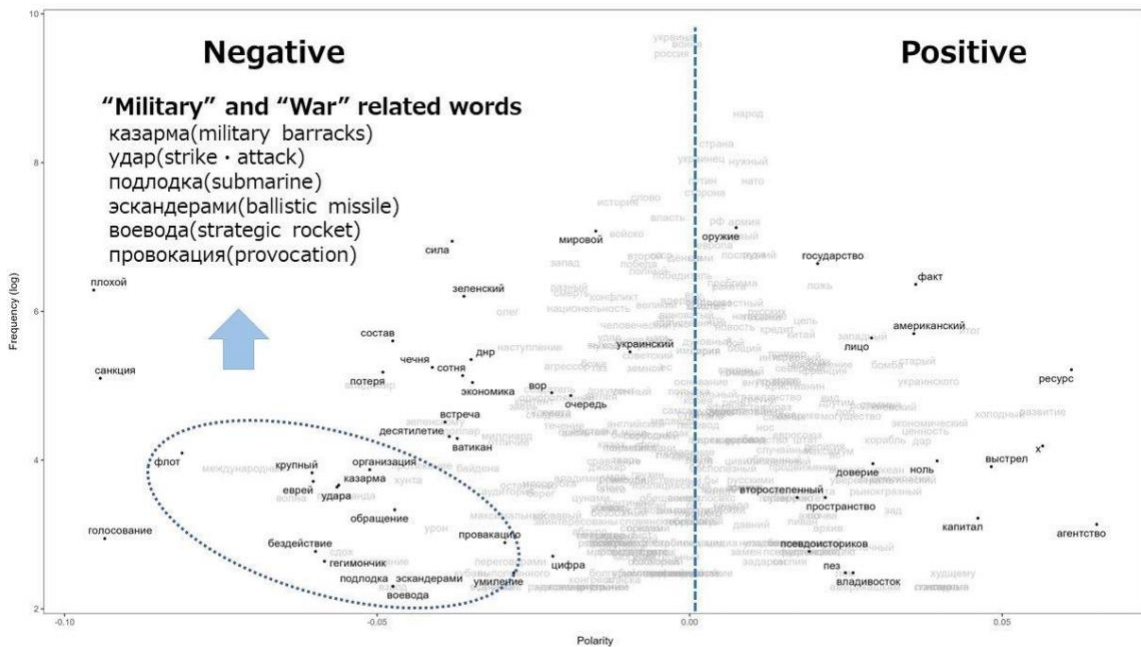


Figure 2

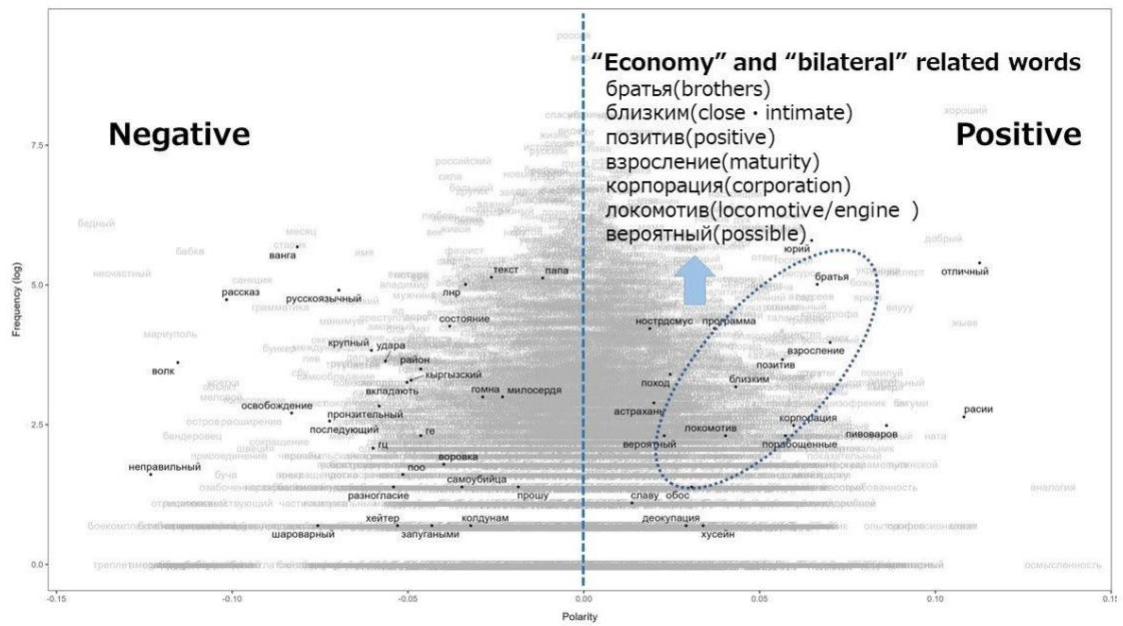


Figure 3

An approach to a corpus-assisted metadiscourse analysis

Isabelle Suremann (Zürcher Hochschule für Angewandte Wissenschaften, Switzerland)

Keywords: corpus-assisted, comparative discourse analysis, media discourse, metadiscourse, COVID-19 vaccination

This poster presents first findings of an exploratory corpus-assisted comparative discourse analysis, which is part of an ongoing dissertation project. This analysis examines metadiscursive statements about media discourse on COVID-19 vaccination, comparing data from two regions with similar measures and restrictions to combat the COVID-19 pandemic but significantly different vaccination rates: the German-speaking part of Switzerland and the Republic of Ireland (Statista 2022).

As public discourse shapes the actions and consciousness of individuals (Spitzmüller/Warnke 2011), it can be assumed that media discourse on the COVID-19 vaccination significantly influenced individuals' decisions for or against vaccination. It can be further assumed that the analysis of metadiscursive statements helps to shed light on individuals' responses to media discourse. In this analysis, metadiscursive statements are defined as statements about media discourse, e.g., discussions about semantics, problematisation of keywords, claims of absences and/or silencing and statements about the order of discourse. As in every discourse analysis, features such as intertextuality and recurrence are crucial (Schröter 2021).

To conduct this analysis, two corpora were compiled: the Swiss corpus, consisting of 3,923 readers' comments on articles on the COVID-19 vaccine from five different Swiss newspapers with a total of 148,516 tokens and the Irish corpus, consisting of 3,861 Facebook comments on the COVID-19 vaccine from five different Irish newspapers with a total of 102,440 tokens. The analysis will follow an iterative process (Dreesen & Judkowiak 2011) and incorporate both corpus-driven and corpus-based methods (Bubenhofer 2009). First, data will be collected using topic modelling, n-gram and keyword analysis. Second, hypotheses based on these data will be generated and tested by applying them to the corpus material.

The aim of this analysis is of exploratory nature. It is assumed that based on readers' and Facebook comments, prototypical metadiscursive statements can be formulated, which can later be applied to larger and broader corpus material, e.g., media and political discourse.

References

- Bubenhof, Noah. (2009). *Sprachgebrauchsmuster: Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. 1(4). Germany. De Gruyter, Print.
- Dreesen, Philipp and Judkowiak, Joanna. (2011). Passiv im Osten, kollektiv schuldig und selbstverständlich in Europa – Kritik an deutschen und polnischen Schulbüchern des Faches Geschichte mittels kontrastiver Diskurslinguistik. *Aptum. Zeitschrift für Sprachkritik und Sprachkultur*. 7(1): 1–31.
- Schröter, Melani. (2021). Diskurs als begrenzter Raum. Metadiskurs über den öffentlichen Diskurs in den neurechten Periodika Junge Freiheit und Sezession. In *Skandalisieren, stereotypisieren, normalisieren. Diskurspraktiken der Neuen Rechten aus sprach- und literaturwissenschaftlicher Perspektive*, herausgegeben von Steffen Pappert, Corinna Schlicht, Melani Schröter, Stefan Hermes. Hamburg. Buske, 51-66.
- Spitzmüller, Jürgen and Ingo H. Warnke (2011). In *Diskurslinguistik. Eine Einführung in Theorien und Methoden der transtextuellen Sprachanalyse*. Berlin. De Gruyter.
- Statista. (Stand: 20. Dezember 2022), Bevölkerungsanteil mit COVID-19-Impfung nach ausgewählten Ländern weltweit. <https://de.statista.com/statistik/daten/studie/1203308/umfrage/impfstoffabdeckung-derbevoelkerung-gegen-das-coronavirus-nach-laendern/> (Accessed January 10, 2024)

Analysis of the media discourse surrounding the 2021 Polish-Belarusian border crisis

Zuzanna Maria Szutta (Uniwersytet Jagielloński, Poland)

Piotr Topolski (Uniwersytet Jagielloński, Poland)

Keywords: Polish-Belarusian border crisis, corpus discourse analysis, media language, refugees, migration

The paper analyses media narratives in Polish newspapers, magazines and media sites regarding the 2021 refugee crisis on the Polish-Belarusian border. The aim of the research was to test the accuracy of the hypothesis that there are two parallel narratives in the media discourse surrounding this topic: one based on security concerns, emphasising the potential danger coming from the migrants (later referred to as the security-based narrative); and the other, recognising them as victims and appealing to the readers' empathy (later referred to as the human rights-based narrative).

The study analyses the keywords of the 17 most influential Polish online publications, which were selected based on the Institute of Media Monitoring [IMM 2021] ranking for October 2021, when the media coverage of the crisis was at its peak. Articles were selected with the use of curated search engine prompts and then collected automatically through web scraping methods.

The texts were later compiled into one specialised corpus consisting of 1441 articles, which was used during the analysis as a reference corpus for finding the key words in every individual publication's corpus. Because Polish is a highly inflected language, the texts had to be lemmatised before any analysis was possible. This process was conducted thanks to a neural network model for the SpaCy Python library. Keywords were then analysed in context to determine what narrative about the crisis each publication was building.

The study verified the hypothesis that there are two opposing narratives about the migrant crisis at the Polish-Belarusian border in the Polish media. The discourse concerning the crisis was polarised, with there being a slightly larger number of strong supporters for the narrative concerned with national security than those empathising with the migrants. There was very little overlap between the two narratives. No publications focused on the security issues and simultaneously found the humanitarian aspect important enough for the keywords to reflect it. Only traces of overlap could be found in the human rights-oriented publications.

This lack of overlap shows that most publications chose to omit certain aspects of the crisis in ways that focused the blame on specific actors. The security-oriented publications blamed the Belarusian government and Lukashenko, seemingly omitting the questionable actions of the Polish authorities. On the other hand, most human rights-oriented publications ignored the Belarusian involvement, putting the whole blame on

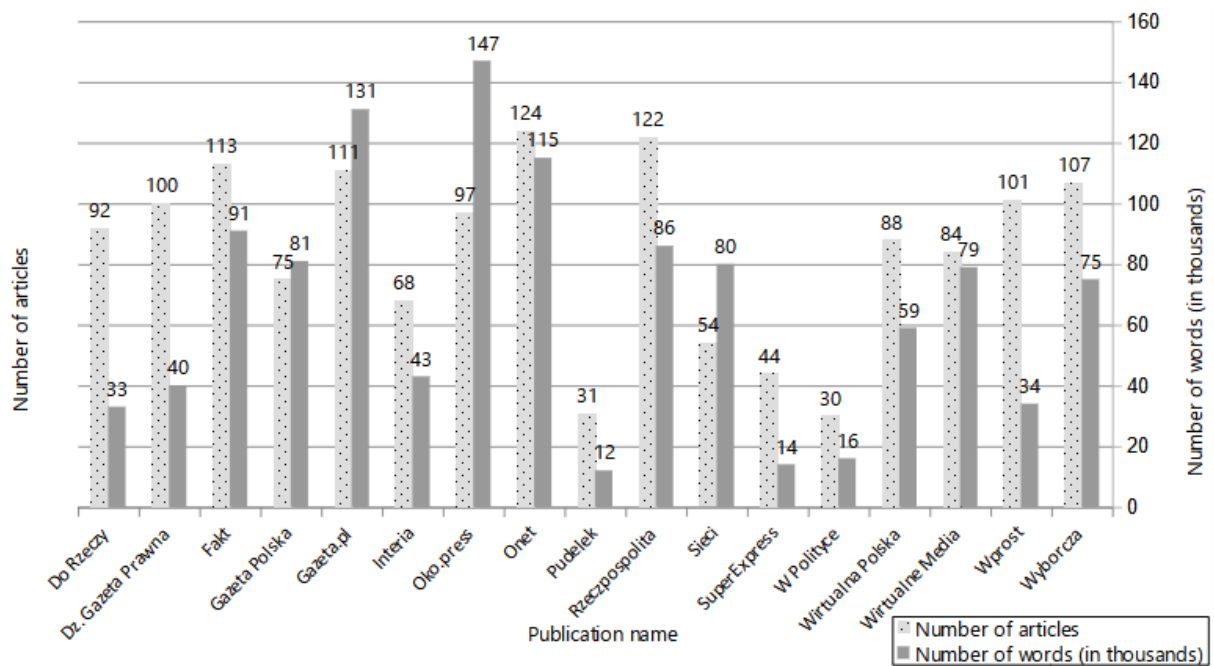
the Polish government. This suggests a severe politicisation of the topic, which was detrimental to the quality of the media coverage.

References

- Anthony Laurance. (2022). AntConc (4.2.0) [computer program], Waseda University, Tokio.
- Anthony Laurance. (2023). *Common Statistics used in corpus linguistics*, available online at https://www.laurenceanthony.net/resources/statistics/common_statistics_used_in_corpus_linguistics.pdf (Accessed: 25 July 2023)
- Baker Paul. (2006). *Using Corpora in Discourse Analysis*. London: Continuum.
- Baker Paul, Anthony McEnery. (2015). *Corpora and discourse studies. Integrating discourse and corpora*. London: Palgrave Macmillan.
- Ba Baziur, Grzegorz. (2022). Operation "Sluice". The so-called migration crisis at the Polish-Belarusian border: an example of hybrid actions taken in the second half of 2021 as documented in the reports of the Polish border guard, *Security. Theory and Practice*, 56(1): 133-150, Krakowska Akademia im. Andrzeja Frycza Modrzewskiego, Kraków.
- Bekić, Janko. (2022). Coercive engineered migrations as a tool of hybrid warfare: a binary comparison of two cases on the external EU border, *Croatian Political Science Review*, 59(2): 141-169, Fakultet političkih znanosti u Zagrebu, Zagreb.
- Berzins, Viesturs. (2022). Hybrid warfare: weaponized migration on the eastern border of the EU?. *The Interdisciplinary Journal of International Studies: Crisis*. 22(1). Aalborg University: Aalborg.
- Biber D. et al., (2007). *Discourse on the move using corpus analysis to describe discourse structure*. Amsterdam: John Benjamins Publishing Company.
- Bloch, Karolina. (2021). Uchodźca, czyli kto? Wizerunek uchodźców na granicy polsko-białoruskiej w internetowych portalach informacyjnych – aspekty prawne i kryminologiczne. *The Polish Bulletin of Criminology* 21: 19-55
- Bodnar, Adam and Agnieszka Grzelak. (2023). The Polish–Belarusian Border Crisis and the (Lack of) European Union Response. *Bialystok Legal Studies* 28(1): 57 - 86.
- Conrad, Susan. (2002). Corpus linguistic approaches for discourse analysis. *Annual Review of Applied Linguistics*, 22: 75-95.
- Dudzińska K., (2021). *The Baltic States' perspective on the forced migration crisis at the EU border with Belarus*, The Polish Institute of International Affairs, available online at <https://www.pism.pl/publications/the-baltic-states-perspective-on-the-forced-migration-crisis-at-the-eu-border-with-belarus> (Accessed: 23 July 2023)
- European Parliament, (26.11.2020). *European Parliament resolution of 26 November 2020 on the continuous violations of human rights in Belarus, in particular the murder of Raman Bandarenka*, Brussels, available online at https://www.europarl.europa.eu/doceo/document/TA-9-2020-0331_EN.html (Accessed: 22.07.2023.)
- Filipec, Ondřej. (2022). Multilevel analysis of the 2021 Poland-Belarus Border Crisis in the Context of Hybrid Threats. *Central European Journal of Politics*. 8(1), 1-18.
- Gilis A., (2023). Lemmatization. In *A guide to artificial intelligence in the enterprise*. Tech Target. <https://www.techtarget.com/searchenterpriseai/definition/lemmatization>
- Gray, Bethany and Douglas Biber. (2011). Corpus approaches to the study of discourse. *Continuum Companion to Discourse Analysis*. London, New York. Continuum. 138-152,
- Grześkowiak, Maciej. The "guardian of the treaties" is no more? The European Commission and the 2021 humanitarian crisis on Poland–Belarus Border. *Refugee Survey Quarterly*, 62(1): 81–102.
- Halemba, Agnieszka. (2022). Ethnographic snapshot: "Europe in the woods: reflections on the situation at the Polish-Belarusian border. *Ethnologia Europaea*. 52(1): 1-13.

- InfoSecurity24, *Atak reżimu Łukaszenki na Polskę i cyniczne wykorzystanie kryzysu afgańskiego [OPINIA]*, 26.08.2021, available online at <https://infosecurity24.pl/atak-rezimu-lukaszenki-na-polske-i-cyniczne-wykorzystanie-kryzysu-afganskiego-opinia> (Accessed: 22.07.2023.)
- IMM, *WP.pl najbardziej opiniotwórczym medium października 2021*, 29.11.2021., available online at https://www.imm.com.pl/wp-pl-najbardziej-opiniotworczym-medium-pazdziernika-2021/?fbclid=IwAR2Q49vcjwW1j180MyEtPjvCzbcM04KWpFFavRYFs2D4XzMNITEa_hwg3uU (Accessed: 21.07.2023.)
- Jas-Koziarkiewicz, Marta. (2023). *The image of migrants during the migrant crisis on the Polish-Belarusian border in 2021 and 2022 in TVP's news service "Wiadomości".* *Studia Politologiczne*. 68: 341-361.
- Jurek, Krzysztof. (2022). *Działalność organizacji pozarządowych w kontekście kryzysu na granicy polsko-białoruskiej (2021-2022).* *Wschodnioznawstwo*. 16: 123-138.
- Klaus, Witold and Monika Szulecka. (2022). *Departing or Being Deported? Poland's Approach towards Humanitarian Migrants.* *Journal of Refugee Studies*. 36(3): 467-488.
- Koźdoń-Dębecka, Monika. (2023). *Polaryzacja medialna na przykładzie kryzysu migracyjnego na granicy polsko-białoruskiej latem 2021 roku w relacjach trzech polskich telewizyjnych serwisów informacyjnych.* In *Media Biznes Kultura*. 14: 161-174.
- Ministerstwo Obrony Narodowej [@MON_GOV_PL], (14.10.2021). *Białoruskie służby zabierają migrantom pieniądze i dokumenty oraz zmuszają ich do nielegalnego przekraczania polskiej granicy* [Tweet]. Twitter. https://twitter.com/MON_GOV_PL/status/1448629108733272069 (Accessed: 22.07.2023.)
- Nylec, Wiktoria. (2023). *Frame analysis of the discourse of Polish weekly magazines on the so-called "migrant crisis" on the Polish-Belarusian border in 2021 based on the example of "Sieci" and "Przegląd".* In *Media Studies/Studia Medioznawcze*. 24(2): 179-189..
- Olbrycht, Paweł. (2022). *Bezpieczeństwo Rzeczypospolitej Polskiej w kontekście sytuacji na granicy polsko-białoruskiej w okresie sierpień-grudzień 2021 r. – możliwości prawne i podjęte działania.* *Polityka i Społeczeństwo*. 20(2): 142-154.
- Polish: spaCy Models Documentation*, available online at <https://spacy.io/models/pl> (Accessed: 21.07.2023.)
- The Week, *Belarus dictator threatens to 'flood EU with drugs and migrants'*. 28.05.2021., available online at <https://www.theweek.co.uk/news/world-news/europe/952979/belarus-dictator-threatens-flood-eu-with-drugs-migrants-avoid-sanctions> (Accessed: 21.07.2023.)
- Tkachuk, Olesia. (2022). *Znaczenie kryzysu migracyjnego na granicy polsko-białoruskiej dla polityki imigracyjnej i azyłowej Unii Europejskiej.* *Wschodnioznawstwo*. 16: 103-122.
- Tóka, Gábor. (2021), *The 2020 Belarusian Presidential Election.* In *Studia Politica. Romanian Political Science Review*. 21(1): 211-230.
- Tymińska Ada. (2022). *Managing fear or rationality? Anti-refugee hate speech on the Polish-Belarusian border.* In *Kwartalnik Trzeci Sektor*. 59-60(3-4): 30-49.
- Wawrzusiszyn andrzej. (2022). *The migration crisis on the Polish-Belarusian border and its influence on the safety of Poland.* *Nowa Polityka Wschodnia* (2). Wydawnictwo Adam Marszałek. Toruń, 45-65
- Wielki Słownik Języka Polskiego. *Uchodźca* <https://wsjp.pl/haslo/podglad/23011/uchodzca>, (Accessed: 22.07.2023.)

Appendix



The number of articles and words from each publication used in the reference corpus