



# RACHELLE VESSEY

**CARLETON UNIVERSITY** 

# FROM CROSS-LINGUISTIC TO INTERSECTIONAL CORPUS-ASSISTED DISCOURSE STUDIES

### **CITATION**

Vessey, R. (2024). From cross-linguistic to intersectional corpus-assisted discourse studies. *Journal of Corpora and Discourse Studies*, 7(1):7–20

### **ABSTRACT**

Ten years ago, I highlighted challenges arising from the application of CADS to multilingual datasets in an approach called 'cross-linguistic corpus-assisted discourse studies' (Vessey, 2013). In the intervening years, the notions of superdiversity and translanguaging have been largely transformative in the fields of applied and sociolinguistics; research applying these notions has raised important questions about boundaries between languages and the nature of diversity in contemporary social contexts (e.g., Blommaert & Rampton, 2011). Drawing and building on these theoretical advances, in this paper I propose to resituate cross-linguistic CADS within a broader intersectional CADS framework (Candelas de la Ossa, 2019; Jaworska & Hunt, 2017; Hunt & Jaworska, 2019; Kitis, Milani, & Levon, 2018; Subtirelu, 2015). Specifically, I underscore the methodological contributions that CADS research can make to the study of intersectionality (Nash, 2008) and I suggest how intersectional theories can support and enrich CADS researchers' arguments about 'non-obvious' meaning (Partington, 2017).

### **KEYWORDS**

### CONTACT

intersectionality, multilingualism, translanguaging, superdiversity Rachelle Vessey, School of Linguistics and Language Studies, Carleton University, School of Linguistics and Language Studies, Carleton University, Ottawa, Ontario, Canada K1S 5B6 RachelleVessey@cunet.carleton.ca

### DOI

ORCID

10.18573/jcads.117

0000-0002-2350-7758

### **ISSUE DOI**

**LICENSE** 

10.18573/jcads.v7

© The authors. Available under the terms of the CC-BY 4.0 license

Manuscript accepted 2024-02-05

# From cross-linguistic to intersectional corpus-assisted discourse studies

### Rachelle Vessey

Carleton University

### 1. Introduction

One of the central appeals of corpus-assisted discourse studies (CADS) is its ability to identify 'non-obvious meaning' (Partington, 2017). Throughout my own research, findings have highlighted that one form of non-obvious meaning emerges from language choice. The use of multiple languages can produce meanings that work on multiple levels and the use of one language rather than another can produce dramatically different discourses (see e.g., Freake, Gentil, & Sheyholislami, 2011; Vessey, 2016). The meanings that can emerge from a multilingual perspective are generally non-obvious, and perhaps as a result they tend to be overlooked. Using multilingualism as a lens through which to identify non-obvious meaning has not been part of a traditional CADS approach, which is why I articulated the challenges emerging from cross-linguistic CADS (Vessey, 2013). As I have focused on and worked with cross-linguistic CADS for many years, I have also identified other factors that function alongside language choice in construing different meanings. In the effort to address these factors using CADS and build on the foundational work of others in this area (e.g., Candelas de la Ossa, 2019; Jaworska and Hunt, 2017; Kitis, Milani, & Levon, 2018; Subtirelu, 2015), I have started to develop an intersectional CADS approach for my research. While the motivation for the earliest intersectional research was to identify and address diverse and interconnecting forms of marginalization (Crenshaw, 1991), intersectionality can also serve as a useful opportunity to align CADS with mainstream and emerging theories in sociolinguistics and beyond. This short paper does not constitute a "how to" for intersectional CADS, but instead highlights some potential advantages that arise from aligning CADS with intersectionality theory, starting with a focus on language.

This paper proceeds with a brief introduction to non-obvious meaning and multilingualism, followed by a short overview of intersectionality. I then provide some suggestions as to how CADS can be used to study intersectionality and, in turn, how intersectionality can enhance our understanding of non-obvious meanings via the consideration of multilingualism within corpora. I conclude by arguing that the intersection between language choice and social phenomena — identifiable through discourse — is perhaps the easiest starting point for beginning to address intersectionality in CADS research.

## 2. Non-obvious meaning and multilingualism

The search for non-obvious meaning using corpus tools is not new (Partington, 2017, p. 339). Stubbs (2009), for example, argues that corpus methods serve as 'estrangement de-

vices' that 'force necessary distance between the observer and the way in which we normally experience running text' (p. 228). It is by changing from our learned way of reading (a) text to a more nonintuitive reading of a corpus (Tognini-Bonnelli, 2010), i.e., through estrangement, that we begin our search for non-obvious meanings. While Baker (2023) describes the focus on non-obvious meanings as the general incorporation of 'what is surprising' (p. 21) into CADS analysis, Partington (2017) proposes more specifically that non-obvious meanings can be categorized into different types or 'varieties' that exist on a cline 'between the pretty obvious to the counter-intuitive' (p. 362). Any single piece of research, he notes, may produce combinations of observations that are more naturally suited to one end of the cline than the other.

For findings to be meaningful at all, the corpus itself must be meaningful. Adhering to Sinclair's (2004) principle will help to ensure this is the case: 'The contents of a corpus should be selected without regard for the language they contain, but according to their communicative function in the community in which they arise'. Adhering to Sinclair's principle in CADS research means that we should not prescribe the linguistic features that should occur in a corpus. As such, a corpus will inevitably include linguistic variation and we must assume that such variation is meaningful (i.e., it has communicative functions in the community). While the field of corpus linguistics has evolved rapidly since 2004, Sinclair's principle nonetheless remains pertinent and is alluded to in a recent introduction to CADS, where Gillings, Mautner and Baker (2023) claim that all CADS projects 'have a social question at their centre rather than a purely linguistic one' (p. 1). Irrespective of the particular social question, naturally-occurring language use in society is increasingly recognized as containing meaningful diversity.

When we analyze language use, we are analyzing choices (Halliday, 2013). While discourse analysis often focuses on the choice of one word over another word, there is also the choice of one language over another language, and such choices are meaningful. Even in ostensibly monolingual contexts, we often have multilingual resources available to us. The work of Hill (e.g., 1995) has revealed that the use of a word or phrase borrowed from another language (e.g., hasta mañana used in English instead of goodbye) has connotative meanings that index stereotypes about the culture where that language is spoken; these indexical meanings are in addition to the denotative meaning(s) of the borrowed words or phrases. Similarly, the work of Kelly-Holmes (2005) has highlighted the role of multilingualism in advertising, where borrowed words, phrases, and even accents are used to connote cultural stereotypes. This is not to say that language choice determines how we see the world (as per a strong form of the Sapir-Whorf hypothesis [Lucy, 1997]), but rather that the choice of one language over another construes particular meanings. For example, it may be that specific topics and issues are discussed more in one language rather than another because culture is reflected in language. Some cultures even have specific ways of viewing and talking about (the role of) language, which Schiffmann (1996) refers to as 'linguistic cultures': 'the set of behaviors, assumptions, cultural forms, prejudices, folk belief systems, attitudes, stereotypes, ways of thinking about language, and religio-historical circumstances associated with a particular language' (p. 5). This means that some cultures may have more self-conscious linguistic traditions, whereas in other cultures, the role of language is much less explicit.

More recently, the notions of superdiversity and translanguaging have emerged and been adopted by applied linguists and sociolinguists to account for the diversities evident in society and communication. 'Superdiversity' refers to a level and kind of complexity surpassing anything previously experienced in a particular society (Blommaert and Rampton, 2011). Translanguaging, broadly speaking, serves as an alternative to structuralist notions pertaining to code-switching, and emerged alongside related terms such as polylanguaging, polylingual languaging, multilanguaging, heteroglossia, hybrid language practices, among many others. Li Wei (2018) frames translanguaging as a 'post multilingualism' concept, where:

having many different languages is no longer sufficient either for the individual or for society as a whole, but multiple ownerships and more complex interweaving of languages and language varieties, and where boundaries between languages, between languages and other communicative means, and the relationship between language and the nation-state are being constantly reassessed, broken, or adjusted by speakers on the ground. Concepts such as native, foreign, indigenous, minority languages are also constantly being reassessed and challenged. What is more, communication in the 21st century requires much more involvement with what has traditionally been viewed as non-linguistic means and urges us to overcome the 'lingua bias' of communication. The Post-Multilingualism era raises fundamental questions about what language is for ordinary men and women in their everyday social interactions. (p. 15)

In other words, the field of linguistics has experienced a poststructuralist turn such that traditional notions of language are being questioned (e.g., Pennycook, 2020).

In the study of obvious and non-obvious meaning alike, the 'post-multilingualism' turn raises questions because corpora are by and large imagined as monolingual entities by corpus linguists. Nurmi and Rütten (20171) explain: 'The monolingual corpus as a monolithic, single-language database, representative of the language of likewise monolingual speakers or writers, is a tacit and probably only half-conscious, but convenient, invention of the corpus linguist' (p. 1). Even in the relatively rare cases where multilingualism is addressed by CADS researchers (see Partington, Duguid, & Taylor, 2013, p. 188; Taylor and Del Fante, 2020), research has tended to rely on corpora that are conceived of as subdivided by language (Marchi and Taylor, 2018, p. 9). Mixed language corpora are few and far between (see Mair, 2013) and it is rare for corpus linguists to adopt terms like translanguaging. Searches of the journals International Journal of Corpus Linguistics, Corpora, and The Journal of Corpora and Discourse Studies reveal only a single research article where the term 'translanguaging' is in use (Kim, 2023); there are no instances of the term 'superdiversity'. Moreover, a superdiverse and translanguaging approach involves more than the cross-linguistic research (Vessey, 2013). Thus, despite the aforementioned growing body of work in applied and sociolinguistics that has challenged the existence of language boundaries, the question remains how to address multilingual diversity in a corpus both in conceptual and practical terms.

The practical challenges are especially pervasive because of the role of frequency as a primary way of establishing meaning within a corpus (see e.g., Baker, 2023, p. 80). Frequency — and frequency-based patterns more generally — underpin most corpus techniques, telling us 'which items and clusters are common in a particular set of texts of a certain discourse' (Partington and Marchi, 2015, p. 217). The frequency-based approach raises issues for the study of multilingualism in CADS research because in naturally-occurring discourse a range of different languages will rarely be used to the same extent (nor should they be). Language mixing does not require equivalent contributions from different languages. Even single instances of multilingualism can reveal particular nonobvious meanings (a point raised by Stubbs in 1998; see e.g., Jaworska and Leuschner, 2018). Given the dominance of different languages in a mixed language corpus, a frequency-based approach will highlight items from the dominant language as most frequent. Most words and phrases from the lesser-used language(s) will appear less prominently and, in some cases, could be overlooked in the analysis, thereby perpetuating the imagined monolingualism within the dataset. Of course, good practice tells us that both high and low frequency items must be examined within a corpus (e.g., Baker and Heritage, 2022; Stubbs, 2001). In their overview of CADS research, Narty and Mwinlaaru (2019) highlight how a focus on frequent usages and repeated patterns leads to overlooking the impact of 'outstanding singular texts' (p. 208). Further, they highlight that a focus on frequency can lead to overlooking what could have been said but was not (Narty and Mwinlaaru, 2019, p. 208). Here, Partington's (2014) typology of absences is instrumental in allowing researchers to account for non-obvious meanings via a range of low frequency phenomena in corpora.

Thus, while existing tools, techniques, and typologies can address multilingualism within corpus-assisted discourse studies, these only go so far in addressing the meaning(s) that multilingualism can have. This is especially compounded because of the predominance of the English language in corpus linguistics. As I argue in this paper, starting with the recognition that language choice is a meaning-making device, we can then identify how it works in tandem with social phenomena under discussion within the discourse. By looking at the intersections of how meanings are made, we can obtain greater insight into the non-obvious meanings in a corpus.

### 3. Intersectionality

Intersectionality emerged in the late 1980s and early 1990s from critical race studies (Crenshaw, 1991) as a way of challenging categorization on the basis of single identity groups. Intersectionality encourages us to consider that all individuals are members of multiple social groups; this multiple membership leads to the (in)accessibility of particular resources and, in turn, the multiple subordinations that are experienced by some individuals. Intersectionality has raised awareness of the diversity within each social category, highlighting how social categories overlap and at times are closely intertwined. Together, these categories can serve to effect marginalization and oppression, as certain aspects of

social memberships become foregrounded or backgrounded, resulting in disadvantage or privilege. While some scholars contend that intersectionality should be reserved exclusively for the study of multiple marginalizations, rather than any nexus of identities, intersectionality has been found to be useful well beyond the original focus on race and gender (Kitis *et al.*, 2018, p. 152; Nash, 2008). I adopt a broader understanding of intersectionality where many different concerns — including language — can be captured within the remit of intersectionality. With a more inclusive remit, I contend that we can learn more about the multiple and intersecting forms through which oppression occurs.

The importance of intersectionality is that it provides a theoretical framework and impetus to capture not only compartmentalized singular or binary phenomena, but also multiplicities of these. As Levon (2015) has argued: 'If it is the goal of our research to understand how social forces inform and constrain observed practice, then it is incumbent upon us to place this intersectional complexity at the heart of our analyses. In practice, this is achieved by working to identify the multiplicity of categories, ideologies, and forces that undergird any observed social phenomenon' (p. 297). Still, researchers have found that despite its theoretical utility, intersectionality remains difficult to execute in practice (McCall, 2005, p. 1772) because of the lack of a rigorous method for examining multiple subject positions (Nash, 2008, p. 89–90). As Chang and Culp (2002) have asked, 'How does one pay attention to the points of intersection? How many intersections are there? Is the idea of an intersection the right analogy?' (p. 485).

It is here that I propose that CADS has much to offer, in particular through the lens of non-obvious meanings. The contribution of CADS also seems a natural extension of existing work in the field, including (but not limited to) cross-linguistic analysis. One core way through which CADS researchers derive non-obvious meanings is via comparison. In fact, Partington (2008) has described CADS as 'inherently comparative' (p. 96). One particular motivation for comparison is the 'corroboration impulse' (Partington, 2017), which has been discussed by Marchi and Taylor (2009) as well as in Jaworska and Kinloch's (2018) and Baker's (2020) discussions of triangulation. However, such comparisons can become problematic in that they might produce or reproduce binaries and categories (see Vessey, 2013). For example, in their research on judicial narratives surrounding women who kill, Potts and Weare (2018) argue that direct comparison (e.g., of women who kill versus men who kill) can serve to "other" social actors against one another. Another potential obstacle is the tendency in CADS to focus on frequency-based findings. Crenshaw (1991) herself has alluded to the problem of frequency-based dominance in her lamentation that 'the political demands of millions speak more powerfully than the pleas of a few isolated voices' (p. 1241). When frequency is used to generate categories, these can become problematic in that they might conflate or ignore intragroup differences (Crenshaw, 1991, p. 1242). This is something that can arise in CADS research, where '[f]avouring dominant and established categories in analysis can essentialise and reify differences' (Vessey, 2013, p. 13) (see also Baker and Heritage, 2022; Hunt and Jaworska, 2019).

Nevertheless, the work of other CADS researchers has highlighted that comparison, corroboration, and triangulation can also be used to identify differences, similarities (e.g., Taylor, 2013), absences (e.g., Partington, 2014) as well as multiplicities. Jaworska and Kinloch (2018) explain that triangulation can be used to 'uncover the playgrounds of ideologies and help understand the mechanics of ideological work in and through discourse' (p. 1113). Similarly, Marchi and Taylor (2018) explain that mixing methods and triangulation hold 'creative power' because 'they allow the researcher to look onto the data from many different windows [and] they account for complexity and help us in dealing with it' (p. 6). Also, CADS research can be used flexibly to generate new ways of understanding the data: 'new categories can emerge bottom-up from our repeated encounters with a specific pattern' (Marchi and Taylor, 2018, p. 5).

If we embark on CADS research with the recognition that 'categories have meaning and consequences' (Crenshaw, 1991, p. 1297), then an intersectional approach means addressing the role of categories in analysis. Specifically, McCall (2005) proposes three specific approaches through which categorical complexity can be conceived of within an intersectional approach: anticategorical complexity, intercategorical complexity, and intracategorical complexity. The first involves the deconstruction of analytical categories using grounded theory and emergent categories. The second approach is where scholars provisionally adopt existing analytical categories in order to explicate the relationships between already constituted social groups and test the utility of the pre-existing categories. Finally, the third approach interrogates boundary-making and boundary-defining process, with a focus on neglected points of the intersection. All of these approaches can draw on CADS tools and techniques. For example, the anticategorical approach essentially speaks to a neo-Firthian approach whereby external theories are not imposed on the data and patterns are allowed to emerge from the corpus (see McEnery and Hardie, 2012, pp. 122-166). Baker (2023, p. 18) suggests that creating categories — rather than deconstructing these — is more familiar to corpus researchers, and this can work well to allow the data to generate 'bottom-up' categories, a practice often used by CADS researchers (Marchi and Taylor, 2018, p. 5). The intercategorical approach is, broadly speaking, that which is adopted by a majority of CADS researchers who undertake comparative research (see Taylor and Heritage, forthcoming). Baker and Levon (2016) adopt such an approach in their analysis of representations of racialized and classed masculinities in the UK print media, triangulating methods to capture both 'frequently articulated representations of masculinity in the corpus and [...] more latent and socially nuanced ideological patterns' (Baker and Levon, 2016, p. 111). Finally, the intracategorical approach can involve the manipulation of corpora themselves into different subcorpora (and sub-subcorpora) to try to address a 'kaleidoscope' (Kenny, 2011, cited in Baker, 2023) of different intersections within the data, allowing us to see which patterns come into focus and recede again as others take their place. Alternatively, tagging and annotation can be used to address intersections of sociodemographic identities within a corpus (see discussion Baker and Brookes, 2022).

Put more simply and explicitly, combining corpus and discourse methods can help us to identify inconsistencies and multiplicities in the meanings of discourse, although there remains work to be done in this area. Notably, there are limitations to the extent to which any methodological approach can fully embrace intersectionality. As Candelas de la Ossa (2019) notes, simply listing multiply marginalized experiences is insufficient, and Baker and Brookes (2022) observe that it is not feasible to carry out comparisons of all possible intersections. While a consideration of combinations of social categories can result in numbers that are too small to meaningfully analyze with corpus methods, this is where qualitative methods and discourse analysis can be used to supplement the corpus findings to help shed light on potentially meaningful intersections within the data (see Baker and Heritage, 2022). As Taylor (2018) has suggested, attention to what we are not looking for can serve as a valuable check on researcher bias.

While CADS has much to offer intersectionality researchers in terms of methodology, likewise intersectionality has much to offer to CADS researchers. For example, intersectionality provides an influential sociotheoretical impetus for the study of absence (Partington, 2014) and similarity (Taylor, 2013). Also, intersectionality might offer CADS researchers ways to grapple with poststructuralist approaches to linguistics, including the notions of superdiversity and translanguaging. Given that corpus methods are increasingly taken-up in linguistic studies of social identity (e.g., Baker and Brookes, 2022), the consideration of intersectionality seems especially pertinent. While Baker (2023, p. 10) has questioned the extent to which corpus research — with its initial emphasis on comparing differences through counting — can be compatible with post-structuralist thinking, if CADS is to produce meaningful findings in the field of linguistics, it will need to address the concepts and contexts that are dominating the field, including intersectionality (Levon, 2015). Indeed, some CADS researchers have been adopting intersectional approaches. For instance, Jaworska and Hunt (2017) examine gender representations in the British press and uncover subtle intersections of gender representations with nationalism and ethnicity. Subtirelu (2015) has also examined the intersection of race and language in student evaluations of lecturers on a 'rate my professor' website. Kitis, Milani, and Levon (2018) have examined representations of Black middle class in Anglophone South African newspapers and, more recently, Candelas de la Ossa (2019) used CADS to identify the multiple marginalizations of survivors of domestic abuse.

At this point, it may not be clear the role of language — and multilingualism — in intersectionality. However, Crenshaw (1991, p. 1249) herself has highlighted that language is one of the ways through which discrimination occurs. Language can serve as a direct or indirect impediment to equality, thereby serving to marginalize individuals on the basis of the language(s) they speak and/or the language(s) they do not speak. This underscores the indexical role of language that extends beyond its purely denotative meaning, as discussed earlier in this paper. It seems that one of the most accessible ways for CADS researchers to begin to address intersectionality is to interrogate the languages used in their corpora because these carry non-obvious meanings which might intersect with other meanings within the corpus. Language use is not always neatly compartmentalized

into separate sections of a corpus or within parallel corpora; language mixing makes the direct comparison of data across languages (e.g., using cross-linguistic CADS) more difficult, if not impossible. Addressing the intersectionality of a corpus via multilingualism starts with an explicit recognition of the variation within a corpus and a presupposition of diversity rather than homogeneity. To begin with, we can simply engage with basic questions proposed by Nurmi and Rütten (2017): 'Is multilingualism reflected in our corpora? If it is, how? And how do we as corpus linguists deal with it?' (p. 1). More generally, what is the (non-obvious) meaning of findings if only one language — like English — is identified and examined? And in particular, what non-obvious meanings are we missing if we filter out non-English data from our corpus altogether? The exclusion of multilingual data is a common practice adopted in CADS, which allows researchers to retain the imagination of a monolingual corpus. However, filtering out multilingualism means that we are using language (and more specifically monolingualism) rather than communicative functions as the basis of our corpus construction. I propose that we will have more access to non-obvious meanings if we adhere to Sinclair's principle (2004) and use the communicative function of language in society to guide our corpus construction. Crucially, we must learn from the work of other researchers and recognize that the communicative functions of language in society are achieved in part through multilingual (translanguaging) features. Ultimately, the aim is to avoid the pitfalls of unicultural and monolingual discourse analysis (cf. Hunt and Jaworska, 2019), which can lead to unwarranted overgeneralizations and the oversight of meaningful intersections.

In other words, if some topics are discussed more in one language rather than another, then the language used in representations of identities, biases, experiences, etc., is part of the intersection. For example, in my own research (e.g., Vessey, 2016) I have found that Canadian identities are expressed through both official languages (and presumably in other languages, too), but the language used is part of the identity marker, even if that tends to not be highlighted explicitly (i.e., metadiscursively). Specifically, English Canadian identity is discursively constructed not only through what is said (i.e., the patterns in the discourse), but also through the use of the English language, which tends to not be mentioned by English Canadians. English Canadians, in fact, tend to discuss French more than English. In contrast, French-speaking Canadians use the French language but also discuss the French language; this language is also used to discursively construct different identities (e.g., Québécois identity). At the same time, a multilingual lens helps to explain the function of borrowed words, like the use of the French term Québécois or Quebecois in English texts, even where an English translation equivalent (e.g., Quebecker or Quebecer) exists. These words are meaningful because of their dual indexicality, in that they not only denote an identity category founded on geography (i.e., the province of Quebec) but also connote an identity category based in language (i.e., French). Here, I contend that the dual indexicality is a site of intersectionality, where a multilingual lens helps to explain an identity category.

Thus, in my own work, I have identified intersectionality as a natural extension of my attention to multilingualism within corpora, and it raises questions about the role of the

English language in CADS research. The predominance of the English language risks obviating the meaning-making functions of English, simply because it becomes taken for granted and common sense. In other words, because English is the unmarked language, the significance of using English — rather than another language — can be overlooked. In their recent overview of CADS research, Gillings, Mautner and Baker (2023) explain that a focus on English language data 'should by no means be taken as the only possible and worthwhile perspective' (p. 2). Certainly, there is only so much work that can be done if researchers have limited proficiency in other languages; nevertheless, a critical and valuable first step would be to recentre intersectional experiences (Candelas de la Ossa, 2019, p. 246) and 'ask the other question' (Matsuda, 1991, cited in Levon, 2015). For example, why English? Interrogating the predominance of English in CADS research means examining the choice (and therefore function) of English over and besides other languages. Such an examination might lead us to better understand how discourses can align and diverge according to the language(s) in use.

### 4. Conclusion

A focus on non-obvious meaning is a useful apparatus through which to conceptualize the aims of CADS research, distilling the potential of complex tools and operations applied to the unwieldy nature of discourse. Non-obviousness also serves as a shorthand for understanding the role of intersectionality within corpus data, i.e., those multiple subject positions that aren't necessarily the first or most intuitive findings that emerge from frequency-based analysis. As Hannem and Schneider (2022) explain:

our views of the world are limited by our positionality. What we see and experience — and what we fail to see and experience — are conditioned by our embodied gender, sexual orientation, skin colour ableness, religion, and culture, but also by the socio-historical lenses and the discourses and language available to us to interpret and speak those experiences. (pp. 4–5)

In order to truly identify non-obvious meaning, then, I suggest that we critically engage with the role of the language that dominates CADS research — English. In other words, the easiest starting point for beginning to address intersectionality in CADS research is to consider the significance of language choice, and then the range of social phenomena that are discussed through that language. This is not to say that intersectionality will be relevant to all CADS research, nor should intersectionality be "shoehorned" into all research; instead, I am proposing that intersectional angles can add to our ability to address non-obvious meanings within a corpus. More specifically, I caution that ignoring the role of English in CADS research may inadvertently exaggerate the voices and/or representations of people who are multiply advantaged, thereby inadvertently marginalizing the voices and/or representations of people who are multiply marginalized (Crenshaw, 1991). While the space constraints of this paper means that it is lamentably short on suggestions for employing intersectional strategies in practice, I maintain that awareness and inten-

tionality are crucial precursors to the meaningful adoption of intersectional perspectives that can contribute to CADS research on non-obvious meaning.

### Competing interests

The author has no competing interests to declare.

### References

- Baker, P. (2020). Analysing representations of obesity in the Daily Mail via corpus and down-sampling methods. In P. Baker & J. Egbert (Eds.), *Using corpus methods to triangulate linguistic analysis* (pp. 85–108). London: Routledge. https://doi.org/10.4324/9781315112466-4
- Baker, P. (2023). *Using corpora in discourse analysis* (2nd ed.). London: Bloomsbury. https://doi.org/10.5040/9781350083783
- Baker, P. & Brookes, G. (2022). Analysing language, sex and age in a corpus of patient feedback: A comparison of approaches. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009031042
- Baker, P. & Heritage, F. (2022). How to use corpus linguistics in sociolinguistics: A case study of modal verb use, age and change over time. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (2nd ed.) (pp. 562–575). London: Routledge. https://doi.org/10.4324/9780367076399-39
- Baker, P. &Levon, E. (2016). 'That's what I call a man': Representations of racialised and classed masculinities in the UK print media. *Gender and Language, 10*(1), 106–139. https://doi.org/10.1558/genl.v10i1.25401
- Blommaert, J. & Rampton, B. (2011). Language and Superdiversity. Diversities, 13(2), 1–21.
- Candelas de la Ossa, A. C. (2019). Exceptionalising intersectionality: A corpus study of implied readership in guidance for survivors of domestic abuse. *Gender and Language*, 13(2), 224–250. https://doi.org/10.1558/genl.35094
- Chang, R.S. & Culp, J. M. (2002). After intersectionality. *University of Missouri-Kansas City Law Review*, 71, 485-491.
- Crenshaw, K. (1991). Mapping the margins: intersectionality, identity politics and violence against women of color. *Stanford Law Review, 43*(6), 1241–1299. https://doi.org/10.2307/1229039
- Freake, R., Gentil, G., & Sheyholislami, J. (2011). A bilingual corpus-assisted discourse study of the construction of nationhood and belonging in Quebec. *Discourse & Society,* 22(1), 21–47. https://doi.org/10.1177/0957926510382842

- Gillings, M., Mautner, G. & Baker, P. (2023). *Corpus-assisted discourse studies*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781009168144
- Halliday, M. A. K. (2013). Meaning as choice. In L. Fontaine, T. Bartlett & G. O'Grady (Eds.), Systemic Functional Linguistics: Exploring Choice (pp. 15–36). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139583077.003 PMid:23497379 PMCid:PMC3605346
- Hannem, S. & Schneider, C. J. (2022). *Defining sexual misconduct: Power, media, and #metoo*. Regina: University of Regina Press. https://doi.org/10.1515/9780889778689
- Hill, J. H. (1995). Junk Spanish, covert racism, and the (leaky) boundary between public and private spheres. *Pragmatics*, *5*, 197–212. https://doi.org/10.1075/prag.5.2.07hil
- Hunt, S. & Jaworska, S. (2019). Intersections of nationality, gender, race and crime in news reporting: The case of Oscar Pistorius Olympian and murderer. *Discourse, Context & Media, 30,* 100299, 1–11. https://doi.org/10.1016/j.dcm.2019.100299
- Jaworska, S. & Hunt, S. (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. https://doi.org/10.1558/genl.28858
- Jaworska, S. & Kinloch, K. (2018). Using multiple datasets. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 110–129). London: Routledge. https://doi.org/10.4324/9781315179346-6
- Jaworska, J. & Leuschner, T. (2018) Crossing languages crossing discourses: A corpusassisted study of Kulturkampf in German, Polish and English. *Pragmatics and Society*, 9(1), 117–147. https://doi.org/10.1075/ps.16028.jaw
- Kelly-Holmes, H. (2005). *Advertising as multilingual communication*. London: Palgrave. https://doi.org/10.1057/9780230503014
- Kenny, D. (2011). Lexis and Creativity in Translation: A Corpus Based Approach. Routledge. https://doi.org/10.4324/9781315759968
- Kim, E. Y. J. (2023). A corpus-based study of anglicized neologisms in Korea: A diachronic approach to Korean and English word pairs. *International Journal of Corpus Linguistics*, 28(2), 125–143. https://doi.org/10.1075/ijcl.20055.kim
- Kitis, E. D., Milani, T. M. & Levon, E. (2018). 'Black diamonds', 'clever blacks' and other metaphors: Constructing the black middle class in contemporary South African print media. *Discourse & Communication*, *12*(2), 149–170. https://doi.org/10.1177/1750481317745750

- Levon, L. (2015). Integrating intersectionality in language, gender, and sexuality research. Language and Linguistics Compass, 9(7), 295–308. https://doi.org/10.1111/lnc3.12147
- Li, W. (2018). Translanguaging as a practical theory of language. *Applied Linguistics, 39*(1), 9–30. https://doi.org/10.1093/applin/amx039
- Lucy, J. (1997). Linguistic Relativity. Annual Review of Anthropology, 26, 291-313.
- Mair, C. (2013). World Englishes and Corpora. In M. Filppula, J. Klemola & D. Sharma (Eds.), *The Oxford Handbook of World Englishes* (pp. 103–122). Oxford: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199777716.013.008
- Marchi, A. & Taylor, C. (2009). If on a winter's night two researchers... A challenge to assumptions of soundness of interpretation. *Critical Approaches to Discourse Analysis across Disciplines*, 3(1), 1–20.
- Marchi, A. & Taylor, C. (2018). Introduction: Partiality and reflexivity. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 1–15). London: Routledge. https://doi.org/10.4324/9781315179346-1
- Matsuda, M. (1991). Beside my sister, facing the enemy: legal theory out of coalition. *Stanford Law Review, 43*(6), 1183–1192. https://doi.org/10.2307/1229035
- McCall, L. (2005). The complexity of intersectionality. *Signs, 30*(3), 1771–1800. https://doi.org/10.1086/426800
- McEnery, T. & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511981395
- Narty, M. & Mwinlaaru, I. N. (2019). Towards a decade of synergising corpus linguistics and critical discourse analysis: A meta-analysis. *Corpora*, *14*(2), 203–235. https://doi.org/10.3366/cor.2019.0169
- Nash, J. C. (2008). Re-thinking intersectionality. *Feminist Review, 89*, 1-15. https://doi.org/10.1057/fr.2008.4
- Nurmi, A. & Rütten, T. (2017). How many languages are there in a monolingual corpus? In A. Nurmi, T. Rütten & P. Pahta (Eds.), *Challenging the Myth of Monolingual Corpora* (pp. 1-15). Leiden: BRILL. https://doi.org/10.1163/9789004276697\_002
- Partington, A. (2008). Teasing at the white house: A corpus-assisted study of face work in performing and responding to teases. *Text & Talk, 28*(6), 771–792. https://doi.org/10.1515/TEXT.2008.039
- Partington, A. & Marchi, A. (2015). Using corpora in discourse analysis. In D. Biber & R. Reppen (Eds.), The Cambridge Handbook of English Corpus Linguistics (pp. 216–

- 234). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139764377.013
- Partington, A., Duguid, A. & Taylor, C. (2013). Patterns and meanings in discourse: Theory and practice in corpus-assisted discourse studies. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.55
- Partington, A. (2017). Varieties of non-obvious meaning in CL and CADS: from 'hindsight post-dictability' to sweet serendipity. *Corpora, 12*(3), 339-367. https://doi.org/10.3366/cor.2017.0124
- Partington, A. (2014). Mind the gaps: The role of corpus linguistics in researching absences. *International Journal of Corpus Linguistics*, 19(1), 118–146. https://doi.org/10.1075/ijcl.19.1.05par
- Pennycook, A. (2020). Pushing the ontological boundaries of English. In C. J. Hall & R. Wicaksono (Eds.), *Ontologies of English* (pp. 355–367). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108685153.018
- Potts, A., & Weare, S. (2018). Mother, monster, Mrs, I: A critical evaluation of gendered naming strategies in English sentencing remarks of women who kill. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 31*(1), 21–52. https://doi.org/10.1007/s11196-017-9523-z
- Schiffman, H. (1996). *Linguistic Culture and Language Policy*. London: Routledge. https://doi.org/10.4324/9780203273487
- Sinclair, J. (2004). Corpus and text: Basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxford: Oxbow Books.
- Stubbs, S. (2009). Memorial article: John Sinclair (1933–2007): The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics, 30*(1), 115–137. https://doi.org/10.1093/applin/amn052
- Stubbs, M. (1998). German loanwords and cultural stereotypes. *English Today, 53*(14), 19–26. https://doi.org/10.1017/S0266078400000675
- Stubbs, M. (2001). Words and phrases: Corpus studies of lexical semantics. Oxford: Blackwell.
- Subtirelu, N. (2015). "She does have an accent but...": Race and language ideology in students' evaluations of mathematics instructors on RateMyProfessors.com. *Language in Society, 44* (1), 35–62. https://doi.org/10.1017/S0047404514000736
- Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora*, 8(1), 81–113. https://doi.org/10.3366/cor.2013.0035

- Taylor, C. (2018). Similarity. In C. Taylor & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 19–37). London: Routledge. https://doi.org/10.1177/0075424207311247
- Taylor, C. & Del Fante, D. (2020). Comparing across languages in corpus and discourse analysis: Some issues and approaches. *Meta: Journal des traducteurs/Translators' Journal*, 65(1), 29–50. https://doi.org/10.7202/1073635ar
- Taylor, C. & Heritage, F. (Eds). (Forthcoming). *Analysing Representation: A corpus and discourse textbook*. London: Routledge.
- Tognini-Bonelli, E. (2010) Theoretical overview of the evolution of corpus linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 14–28). London: Routledge. https://doi.org/10.4324/9780203856949-2
- Vessey, R. (2013). Challenges in cross-linguistic corpus-assisted discourse studies. *Corpora,* 8(1), 1–26. https://doi.org/10.3366/cor.2013.0032
- Vessey, R. (2016). Language and media in Canada: Representations, ideologies, policies. London: Palgrave. https://doi.org/10.1057/978-1-137-53001-1\_2