Evaluating corpus & discourse work

Chapter 14. Evaluating corpus & discourse work

Anna Marchi

https://orcid.org/0000-0003-1284-6356

This chapter considers how we evaluate the quality of corpus-assisted discourse analyses. I will seek to convey the perspective you may anticipate from a supervisor, an examiner, or a reviewer: the standard of expectations may change with context, but the overall criteria will remain somewhat consistent at different stages of your academic career. Section 14.1 refers to the overall soundness of research and is concerned with the analysis itself, while section 14.2 discusses the reporting on the analysis. Section 14.3 touches upon the concept of impact, by addressing originality and relevance. Section 14.4 praises constructive criticism and, in part, attempts to redeem Reviewer 2. Finally, section 14.5 is a call for greater awareness and a constant practice of self-reflexivity. Throughout the chapter you will find questions, which you may use as a sort of checklist, and examples of potential pitfalls. A disclaimer: all examples are fictitious and no identification with actual research (published or unpublished) is intended or should be inferred.

14.1. Validity and reliability

As for all scientific endeavours, key criteria in judging the quality of research are *validity* and *reliability*. The two are closely related concepts: validity refers to how well the data correspond to the phenomenon they are meant[1] to represent and how accurately a method measures what it is intended to measure[2], while reliability relates to the consistency of methods and how sure we can be that, using the same data and the same methods, other people would arrive at the same results. So, evaluating validity and reliability of corpus-assisted research implies taking into consideration several layers: the corpus data, the methods employed, and the results obtained.

14.1.1 Evaluating the corpus

Perhaps the most important rule in corpus work is that 'the corpus data we select to explore a research question must be well matched to that research question' (McEnery & Hardie, 2012: 2). This is an extension of two core points:

- ✓ The evaluation of the quality of the corpus: how good is the corpus?
- ✓ The evaluation of the appropriateness of the corpus: what is the corpus good for?

[1] For example, a corpus of *the Guardian* is not representative of the British press, but a corpus sampling newspapers across the political spectrum and covering different kinds of readership might.

[2] For example, that distinctiveness (as measured by Keyword analysis) is the most relevant feature to define a register, a discourse type, or the content of a specific corpus.

Both questions concern representativeness, i.e. whether the texts in the corpus are an adequate representation of the type of language or of the phenomenon under investigation (for a recent examination see Egbert et al. 2022). As John Sinclair famously put it, ultimately 'the results are only as good as the corpus' (Sinclair, 1991: 13). The second point evaluates the corpus with reference to the research question(s): so either the corpus fits the RQ(s) or the RQ(s) have to fit the corpus. We cannot, therefore, make generalisations about language at large based only on a corpus containing texts from a limited variety of language domains. Findings and their interpretation should not (unless with caution) be stretched beyond the boundaries of the corpus itself. For example, an investigation of semantic prosody based solely on newspaper data is likely to be affected by the fact that negativity is the most typical news value (for more on news values see Bednarek & Caple 2017) and therefore negative evaluation is intrinsically prominent in news discourse.

Because CADS work often relies on specialised corpora compiled for specific purposes (Chapter 5), the superordinate point about the general quality of the corpus is crucial. Details about data collection and corpus composition and size make it possible for a reviewer to assess the quality of the corpus and answer the following questions:

- ✓ Are the compilation criteria adequate, with regard to the target(s) of the research?
- ✓ Is the composition homogeneous, with reference to the discourse type(s) under investigation?
- ✓ Is the corpus large enough, with respect to the purpose and the methods?

Often specialised corpora are collected using queries designed to retrieve texts on a specific topic (for example texts about Brexit, or about Covid) from a much larger output. The lexical items implemented in the query must produce a complete and coherent dataset, striking a balance between *precision* and *recall*: 'the query would, ideally, return all relevant articles in the database (i.e. have 100% recall), while avoiding also returning irrelevant articles (i.e. have 100% precision)' (Gabrielatos et al., 2012: 154). In some cases this will be quite straightforward, however, because language is messy and mutable, in other cases we can only expect an approximation. For example: a corpus aimed at studying representations of foreigners containing only texts mentioning the word IMMIGRANT would be blatantly inadequate, since foreignness can be lexicalised in a variety of ways. The question is whether an approximation is good enough or whether it risks compromising the validity of the analysis by systematically skewing the results.

Another factor in determining the adequacy of the corpus with respect to the RQ(s) is the homogeneity of the texts included: *composition variability* (see Gries 2006) may, in fact, considerably affect the analysis and the results. A typical example is, again, newspaper corpora, since the content of newspapers includes very different types of writing and a growing proportion of feature articles. So if, for instance, we are investigating the increasing personalisation of news discourse, we want to make sure that we are actually looking only at news reporting.

But perhaps the most frequent question asked about corpus design is: is the corpus big enough? Size tends to be overrated, in comparison to other characteristics, and there is no one-size-fits-all answer to this preoccupation. As Egbert et al. (2022) point out "large" is a relative concept,

which depends on a number of factors, so the question is a difficult one because it depends on: large enough to do what? If complete representation of the population of relevant texts can be achieved, the corpus should be complete. For example, if you were investigating the official speeches of Donald Trump as President of the United States and only included a selection of speeches in your corpus, this would most likely be deemed a flaw. If the data need – for practical (for example the sheer size of the complete output) or for intrinsic reasons (for instance the unavailability of the complete output) – to be a sample of the texts representing the phenomenon under investigation, the aim is to have a corpus that is proportionate to the generalisations we intend to make, in other words: a valid quantitative analysis may not be obtainable from insufficient data.
Extremely valuable corpus-assisted research has been produced using small corpora and this is due to the fact that small specialised corpora have a key advantage over large corpora: access to context and detailed background information which can contribute to the interpretation of data (see Flowerdew 2008). But perhaps the question about size could be reversed to: is my collection of texts small enough *not* to use corpus methods? Does the size of the dataset and frequency of the target items justify the use of corpus methods and tools, or would a more traditionally qualitative approach be a better fit? A rule of thumb is that if you could have done the same research and obtained the same (or even better) results by reading and analysing the texts sequentially, without employing a concordancer, the analysis may possibly not qualify as properly corpus-assisted/based.

14.1.2 Evaluating the methods

Methodology, according to a study (Ganji & Derakhshan 2020) on criteria for evaluating manuscripts submitted to *Applied Linguistics*, is the criterion that gets the most mentions in the journal's guidelines for reviewers, followed by references, writing quality, and originality. The evaluation of methods seems to correspond to an overall judgement of soundness, but evaluating methods implies taking into consideration a variety of different elements that make up a methodological package. It could be argued that the evaluation of the corpus, discussed earlier, is part and parcel of methods, but if we limit the concept of methods strictly to the analytical steps and tools we can focus on a series of choices along the analytical path. Some prominent decisions are:

- ✓ Choices of segmentation: granularity and identification of units of analysis
- ✓ Choices of and about metrics: statistical measures, spans, cut-offs
- ✓ Choices of operationalisation: identification of lexical items

These choices represent potential pitfalls. The very preliminary choice of units of analysis impacts the entire analysis, therefore the rationale behind the choice should be made clear and explicit. Marchi (2018), for example, discusses how different segmentations of time units deeply affect findings in diachronic analysis, by showing the completely different output of keywords obtained by comparing different time units (for example adjacent months vs adjacent seasons). Similarly the selection of metrics, which affect rankings (see Gabrielatos 2018), and of spans and thresholds should be justified and grounded on methodological reflection, whenever possible. 'Making informed statistical choices is an essential skill for a researcher' (Brezina, 2018: 259); arguments such as "I am going to analyse

the top 100 keywords" or motivations such as "these are the software's default settings" are likely to look rather suspicious.

An aspect that does not receive as much attention as it perhaps ought to is *operationalisation*, that is the definition of lexical items which allow to measure, systematically and as comprehensively as possible, abstract concepts. In CADS, in fact, we are often interested in studying complex phenomena which may have multifaceted discursive realisations, 'we must ensure that the words we choose are valid representatives of the cultural concept in question' (Stefanovitch, 2018: 253). This links back to precision and recall, discussed earlier: the lexical items we look at have to be as plausibly exhaustive as possible. Looking only at a selection of words, without convincingly justifying the selection, is preemptive *cherry-picking*. Cherry picking per se does not make the analysis unsound, but it makes it impossible to determine the soundness of the analysis. I will say more about this when dealing with the evaluation of results in the next section.

In addition to things that can be found to be "wrong" when reviewing research, there are also things that may just be found to be missing. Typically overlooked areas are for example the analysis of similarities (Taylor 2013, 2018) and of absences (Partington 2014).

14.1.3 Evaluating the results

The same layering of criticism may be applied to results: there can be inaccurate findings and incomplete findings. However, it might be difficult to make a clear cut distinction between the two because unrepresentative findings are also unreliable. It all harks back to the aforementioned practice of "cherry picking".

> If a corpus is used as a database of examples, the data can be cherrypicked in order to support different, even contradictory, hypotheses. As Gabrielatos (2013) eloquently puts it, 'corpus linguistics is very easy to do badly'. Hence the importance of a quantitative approach, accounting for all instances and relying on statistics. When we report on the corpus analysis it is essential that we select examples in order to illustrate our interpretations of the data and, in order for the examples to be useful and representative, it makes sense to pick the ones that most clearly evidence the interpretation. Cherry-picking, however, is only fine as long as we pick standard cherries and do not discount counterexamples possibly falsifying our interpretations. (Marchi, 2019: 40)

The "standardness" of the cherries needs to be quantitatively evidenced. The principle of 'total accountability' (Leech, 1992:112 – discussed in Chapter 1) is the very essence of corpus research and CADS must strive for a systematic analysis of all the available evidence. This is of paramount importance because: if you want to find something in a corpus you will! The fact that we are using a corpus is per se no guarantee of greater objectivity or soundness (see Marchi & Taylor 2009). A way to keep in check our biases and our subjective influence is to constantly practice the 'culture of the counterexample' (Partington et al., 2013: 332) and actively pursue contradicting or non-conforming evidence; counterexamples, continue Partington et al. 'though sometimes vexing for the researcher, are an excellent thing for the research' (*ibid.*). Assessing research in this perspective means asking:

- ✓ Is the evidence presented exclusively exemplificatory and corroborative rather than comprehensive? Does the author provide accurate quantification and account for counterexamples?

Other questions concerning the quality of the results, or rather of the interpretation of the results, are to do with the twin and complementary (see O'Halloran & Coffin 2004) dangers of *over-interpretation* and *under-interpretation*. By overinterpretation I mean overgeneralisation of findings beyond corpus boundaries. In fact, if it is true that we attempt to make broader claims about language and society, it is also true that 'findings cannot be generalised beyond the scope of the period and the entity of the material we are handling' (Marchi, 2019: 74). McEnery and Brezina have recently summarised this in their 48 fundamental principles of corpus linguistics:

> Principle 5: By studying corpora, i.e. finite samples of language, we can make general claims about language itself; these claims are probabilistic in nature.
> Principle 5': By studying corpora, i.e. finite samples of language bound by space and time, we can make claims about language itself, with those claims bound by space and time in the same way as our data. (McEnery & Brezina 2022: Appendix 1)

Underinterpretation may come in at least two shapes: lack (or superficiality) of explanation and lack (or superficiality) of linguistic description. The first is the case when the analysis accounts to little more than the reporting of a large number of examples and it can thus be accused of merely being 'running commentary on long quotes'.[i] While the aim (and the attitude) of much CADS research is descriptive rather than explanatory, it is still important not to assume that descriptions will lead to self-evident and unambiguous interpretations and to make interpretations explicit. It is good practice to leave your interpretations coherent but as open as possible, or even suggest competing or overlapping interpretations, and allow the reader to decide.

A more fundamental problem is the lack of attention to language and discourse themselves, which is a criticism often aimed at so-called "bag of words" approaches, where the sole focus is the reporting of statistical and distributional information. '[F]requencies or distributional information cannot, in and of themselves, answer questions of interest; they need to be interpreted by a linguist to provide answers to questions of linguistic relevance' (Larsson et al., 2022: 138). And while quantitative information is fundamental, for all the reasons that we have discussed previously in this chapter, 'the ultimate goal of corpus linguistics' is 'learning about language use with the help of a corpus' (Larsson et al., 2022: 155).
So ask yourself:

- ✓ Does the research merely report findings and examples that are treated as self-explanatory?
- ✓ Does the research look as mere "bean counting" and fails to analyse language and discourse in depth and contextually?

Lack of interpretation leads to one kind of "so what?" reaction, i.e. what can we conclude and why is this interesting? There is then another kind of "so what?", i.e. how is this new?. While ground-breaking findings are a wonderful thing, we need not be overly concerned about impact, after all

'corroboration of what is already strongly suspected is frequently a vital component of scientific advance' (Partington et al., 2013: 9). However, even though knowledge production is more often an incremental evolution rather than a radical leap, we must also avoid 'stating the obvious with an air of discovery' (Marchi, 2019: 51) a.k.a. reinventing the wheel (more on this towards the end of the next section).

14.2. Transparency and thoroughness

When we evaluate a piece of research we are inevitably evaluating not the analysis itself, but the final written trace of the analysis. In other words, a research paper, in the attempt to be coherent to the reader, is the rearrangement in a sequential form of a process that was in practice most likely anything but linear. Analyses tend to be recursive and multifarious: we zoom in, we zoom out, further steps are embedded in the process to pursue new evidence and multiple connections, and so on. This is in the nature of a bottom-up approach, where the patterns that emerge from the data lead the analysis, ever more so of an eclectic approach such as CADS which, despite being primarily inductive, also accommodates, indeed often originates from, knowledge that comes "sideways", for example input from other non-linguistic disciplines, and sometimes even from inferences and speculations on our "real-world" knowledge or beliefs.
Reporting findings is a form of storytelling and an exercise in reducing complexity without compromising transparency, and transparency is the precondition and the manifest expression of accountability. Mirroring the previous section, I will briefly discuss transparency of the data description, of the chronicle of the methods and steps of analysis, and of the presentation of findings:

The level of detail of the data description clearly depends on the aims of the analysis, but a wordcount is hardly ever sufficient. You will need to include information about corpus compilation, indicating for example:

- ✓ Where the texts come from. A corpus of newspaper articles, for example, could be compiled using a database, a news aggregation platform, or the newspaper's website and the origin of the data may have an impact on their format, as well as their content.
- ✓ How the texts were collected (e.g. using a software or a script, or manually) and, in case they needed cleaning-up from boilerplate material (for instance metadescriptions of parts of texts) or duplicates, how they were processed.
- ✓ How the texts are stored, for example. if we collect newspaper output in one file for each article, we can do different things with the data than if we collect the output in one file for each daily edition of the newspaper. Corpus storage is not mere data hoarding, the architecture of the corpus determines its potential uses. The unit of storage identifies also what counts as a text and thus becomes our basic unit of analysis.
- ✓ Relevant contextual information about the sources and the texts. In corpus compilation this is referred to as "corpus enrichment" and it compensates for the 'inevitable loss of contextual information in the making of a corpus […] by means of rich metadata that describe the material' (Ädel et al., 2018: 22).

- ✓ Broader contextual background. An account of the socio-cultural background cannot be comprehensive, but we need to be aware of the assumptions we make about our readership and what knowledge can be taken for granted.
- ✓ If the corpus is annotated (Chapter 6), fine-grained information about the annotation scheme is required.

Once the stage has been set, we need a description of the game (as amply discussed in section 14.1.2), which should include:

- ✓ Clear research questions and aims of the study.
- ✓ Description and justification of methods. For example: in a diachronic analysis (Chapter 12) we will have to explain the whys and the hows of the segmentation of time units.
- ✓ Operationalisation of variables. For example: if we study something like "citizenship" or "populism" as a concept, rather than as a lexeme, we will need to provide a definition and a list (and justification) of the lexical items that describe that abstract concept.
- ✓ Indication and justification of tools and settings, as discussed in the previous section. For example: Why was a tool preferred over others? What statistical measures were chosen and why? What rankings, spans, cut-offs were chosen and why?
- ✓ Description of the main steps of the analysis. Corpus-assisted work is often rather eclectic and cannot (nor should) be reduced to a set of procedures, however an account of the research should document the process in sufficient detail as to make it retraceable and to enable *replicability* (Chalpter 1).[ii]

The visibility of the process is accompanied by the visualisation of the results and, as Gries and Paquot warn, '[i]f good 'Methods' sections help ensure reproducibility, good 'Results' sections ensure comprehensibility and go a long way towards making a reader accept one's conclusions' (Gries & Paquot, 2020: 654). In corpus assisted work we typically deal with two kinds of results: descriptive statistics and elucidatory examples. The matter of "picking standard cherries" has been discussed earlier, but the presentation of quantitative results is another very important element in the evaluation of findings and of the overall analysis. Quantitative findings will tend to be presented through tables and charts which accomplish multiple functions: they are a space-efficient way to report large amounts of information, they grant greater accountability by offering a bird's eye view of results that can then be illustrated by means of examples, they can make part of the process of analysis visible, and visualisation can be an effective way make patterns recognisable. When including graphic representation ask yourself:

- ✓ Does the visualisation add information and value to the analysis or is it redundant?
- ✓ Does the visualisation offer a complete view?
- ✓ How does the visualisation affect perception? For example, does the choice of infographic minimise or exaggerate the phenomenon it represents?[iii]

The write-up of research is a tightrope walk between a thorough, fine-grained and transparent account of the research process, on one side, and a concise and convincing delivery of a research product, on

the other. Borrowing words from a giant, the aim is 'to try to give all of the information to help others to judge the value of your contribution; not just the information that leads to judgement in one particular direction or another' (Feynman, 1974: 12).

14.3. Originality and relevance

When applying for research grants, hyperbolic boasting appears to be a basic requirement. You will be encouraged to present your project in terms of how innovative, ground-breaking, game-changing it is. ERC funding, for example, operates on a high risk / high gain parameter and aims to enable "frontier research" and nothing less than "advances of knowledge to the edge of the unknown".

Innovation and impact are certainly important parameters, however originality can take many shapes and, as mentioned earlier in this chapter, knowledge generation is more often made of small incremental advances than huge leaps. Para-replication of previous research on different data, for example, is an extremely relevant pursuit that is rarely practiced, perhaps because it is perceived as lacking in originality, but which can instead greatly benefit the scope and the depth of our understanding of phenomena (see for example Taylor 2014 or Partington & Zuccato 2018). In the evaluation of a research paper, original work is work that is not merely derivative and the impact of the research is mainly accessed through the discussion of the implications of the results for your hypotheses and the implications for the research field. A thorough review of the literature that demonstrates awareness of the state of the art is the necessary premise for original research, because all new knowledge is, first, of all acknowledgement and the ability to build from what we already know. Missing references to previous work are among the most frequent comments in peer-reviews.

That said, various aspects of originality in your work will be valued, for example: if it develops innovative methods, if it focuses on an underexplored discourse-type or an underrepresented reality, or if it addresses a timely issue.

Timeliness takes us from the dimension of scientific relevance to the more subjective one of social relevance. It is common for CADS research to be concerned with so-called 'areas of pressing concern', as stated on Lancaster's Corpus Approaches to Social Sciences (CASS) website.[iv] Often we are interested in discourses precisely because of the impact they have on the real world and on people's lives, that is because they matter. Engagement with the world we live in, and even commitment to make it a better place, is an element that may add value to your work, but this should not translate into instant research. Research, in fact, is a slow process which cannot chase events, lest timeliness becomes hurriedness.

14.4. Reviewer 2

Along your path you are bound to meet the legendary, despised, and "overmemed" Reviewer 2. Reviewer 2 has become an archetypal character in the academic world and is portrayed as an embittered scholar who delights in harming their peers by dishing out abusive and pointless criticism. While unnecessarily mean feedback is always wrong (and there actually is no such thing as

necessarily mean feedback), sometimes the greater problem is that we, understandably, do not crave criticism. But constructive criticism, and sometimes even disruptive criticism, can substantially improve our work. I would like to share a personal anecdote to support my unpopular claim. I received three reviews on a book proposal, one was very positive but rather short, one was generally positive and very thoughtful, and one was one was extensively negative, hostile, sarcastic and at the time much of it felt just inappropriate. Michael Hoey (a leading corpus linguist best known for his theory of - and 2005 book - *Lexical Priming*), who was editor of the series, told me he agreed that the tone of the feedback was upsetting, but he encouraged me to read past the attitude and make the most of the meticulous comments. His prep-talk went something like this: you won't believe this now, but the negative review is the one that will be most useful to you; once published your work is out there and out of your hands and it will define you, so you want it to be the best you can do. So if someone has devoted time to dislike your work, try to use that to your advantage. Rather than evaluating criticism on the basis of its positivity/negativity, we could try to evaluate it on the basis of its depth of engagement with our work. With distance comes perspective.

So stereotypical Reviewer 2 is bad, but we should be careful not to confuse that with criticism itself being bad. Peer-reviewing is a very important service to the research community and for the advance of disciplinary knowledge. Always try to be honest, kind and constructive when you are evaluating someone else's work, but also remember that the worst kind of feedback is lazy feedback, for instance feedback that is simply dismissive and fails to suggest alternative directions/solutions.

14.5. Self-reflexivity

Good research is research that acknowledges its limitations as well as being able to communicate its strengths. I would like to conclude this chapter with an invitation to practice self-reflexivity. Being critical and inquisitive about how research design, methodological choices, and our own assumptions and biases affect the analysis and the results, increases self-awareness and makes our work more transparent and accountable. Self-reflexivity entails questioning the premises and the process of research and preserving enough distance (and a dose of humbleness and self-doubt) to ask ourselves questions – yet another checklist – such as:

> How has the research question defined and limited what can be 'found'? How has the design of the study and the method of analysis 'constructed' the data and the findings? How could the research question have been investigated differently? To what extent would this have given rise to a different understanding of the phenomenon under investigation? (Nightingale & Cromby, 1999: 228)

Alongside the researcher's impact on the researched we must also acknowledge the intrinsic challenges and constraints of corpus methods (for an extensive discussion see Taylor & Marchi 2018 edited collection), which might help us compensate for them. Corpus-based analyses, for example, are usually monosemiotic and in most cases – since the object of study is the message – they fail to incorporate elements of production and reception in the analysis. On the other hand, precisely because the discipline has matured an interest in self-scrutiny and because of its inbuilt 'exploratory nature' (Partington, 2009: 286), CADS is rather good at binging on context and stepping beyond the corpus to bring in deeper understanding and new ideas. Which brings me to a last bit of evaluation of

research, arguably a more personal one: the papers I find most interesting are those that demonstrate a voracious curiosity for multiple perspectives, exhibit a porous nature, and engage with problems in a creative and holistic way. To seal this with my favourite linguistics' quote: '[a]n open mind is the best guide in linguistics, as in research in general and indeed in life itself' (Johansson, 1991: 6).

References

Ädel, A. (2018). Corpus Compilation. In S. Th. Gries and M. Paquot (Eds.), *A Practical Handbook of Corpus Linguistics* (pp. 3-24). Springer.

Bednarek, M. & Caple, H. (2017). *The Discourse of News Values: How News Organisations Create Newsworthiness*. Oxford University Press.

Brezina, V. 2018. Statistical choices in corpus-based discourse analysis. In C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp. 259-280). Routledge.

Egbert, J., Biber, D., & Gray, B. (2022). *Designing and Evaluating Language Corpora. A Practical Framework for Corpus Representativeness.* Cambridge University Press.

Feynman, R. P. (1974). Cargo Cult Science. Some remarks on science, pseudoscience and learning how not to fool yourself. Caltech's 1974 commencement address. Available online: http://calteches.library.caltech.edu/3043/1/CargoCult.pdf.

Flowerdew, L. (2008). Corpora and Context in Professional Writing. In V. K. Bahtia, J. Flowerdew & R. H. Jones (Eds.), *Advances in Discourse Studies* (pp. 115-131). London: Routledge.

Gabrielatos, C. (2013). Keyword analysis. Presentation given at Dubrovnik Fall School in Linguistic Methods, Centre for Advanced Academic Studies, Dubrovnik, Croatia, 20–26 October. Slides available at: http://repository.edgehill.ac.uk/5932/1/NTNU.Dubrovnik.Keyness.pdf.

Ganji, M., & Derakhshan, A. (2020). Developing a Checklist for Evaluating Research Articles in Applied Linguistics. *Teaching English Language 14*(2), 239-268. DOI:10.22132/TEL.2020.121858

Gries, S. (2006). Exploring variability within and between corpora: some methodological considerations. *Corpora 1*(2), 109-51. DOI: 10.3366/cor.2006.1.2.109.

Gries, S. Th. & Paquot, M. (2020). Writing up a Corpus-Linguistic Paper. In S. Th. Gries and M. Paquot (eds.) *A Practical Handbook of Corpus Linguistics* (pp. 647-659). Springer.

Gabrielatos, C. (2018). Keyness analysis: nature, metrics and techniques. In C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp. 225-258). Routledge.

Gabrielatos, C., McEnery, T., Diggle, P. J., & Baker, P. (2012). The peaks and troughs of corpus-based contextual analysis. *International Journal of Corpus Linguistics*, 37(2), 151-175, DOI: 10.1075/ijcl.17.2.01gab.

Hoey, M. (2005). *Lexical Primingg: A New Theory of Words and Language*. Routledge.

Johansson, S. (1991). Computer corpora in English language research. In S. Johansson & A. Stenström (Eds.), *English Computer Corpora: Selected Papers and Research Guide* (pp.3-6). Mouton de Gruyter.

Larsson, T., Egbert, J., & Biber, D. (2022). On the status of statistical reporting *versus* linguistic description in corpus linguistics: a ten-year perspective. *Corpora* 17(1), 137-157. DOI: 10.3366/cor.2022.0238.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (Ed.), *Directions in Corpus Linguistics: Proceedings of Nobel Symposium 82 Stockholm, 4–8 August 1991* (pp.105-122). Mouton de Gruyter.

Marchi, A. (2018). Dividing up the data: epistemological, methodological and practical impact of diachronic segmentation. In C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp.174-196). Routledge.

Marchi, A. (2019). *Self-Reflexive Journalism. A Corpus Study of Journalistic Culture and Community in the Guardian*. Routledge.

Marchi, A., & Taylor, C. (2009). 'If on a winter night two researchers': A challenge to assumptions of soundness of interpretation. *Critical Approaches to Discourse Analysis across Disciplines 3*(1), 1–20.

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

McEnery, T. & Brezina, V. (2022). *Fundamental Principles of Corpus Linguistics*. Cambridge University Press.

Nightingale, D., & Cromby, J. (1999). *Social Constructionist Psychology*. Open University Press.

O'Halloran, K.A., & Coffin, C. (2004). Checking Overinterpretation and Underinterpretation: Help from Corpora in Critical Linguistics. In C. Coffin, C., Hewings, A., & O'Halloran, K.A. (Eds.), *Applying English Grammar: Functional and Corpus Approaches* (pp.275-297). London: Hodder Arnold.

Partington, A. (2009). Evaluating evaluation and some concluding thoughts on CADS. In J. Morley and P. Bayley (eds.) *Corpus-Assisted Discourse Studies on the Iraq Conflict. Wording the War* (pp. 261-303). Routledge.

Partington, A. (2014). Mind the gaps. The role of corpus linguistics in researching absences. *International Journal of Corpus Linguistics 19*(1), 118-146. DOI: 10.1075/ijcl.19.1.05par.

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

Partington, A., & Zuccato, M. (2018). Europhobes and Europhiles, Eurospats and Eurojibes: Revisiting Britain's EU debate 2000-2016. In A. Čermáková & M. Mahlberg (Eds.), *The Corpus Linguistics Discourse. In honour of Wolfgang Teubert* (pp.95-126). John Benjamins.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocations*. Oxford University Press.

Stefanovitch, A. (2018). *Corpus Linguistics. A Guide to the Methodology*). Language Science Press.

Taylor, C. (2013). Searching for similarity using corpus-assisted discourse studies. *Corpora 8*(2), 81-113. DOI: 10.3366/cor.2013.0035.

Taylor, C. (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.

Taylor, C. (2018). Similarity. In C. Taylor & A. Marchi (Eds.), *Corpus Approaches to Discourse: A Critical Review* (pp. 19-37). Routledge.

Taylor, C. & Marchi, A. (eds.) (2018). *Corpus Approaches to Discourse. A Critical Review*. Routledge.

[i] This is authentic feedback I received from a reviewer (surprisingly Reviewer 1 rather than 2).
[ii] See also 'reconstructability' (Stefanovitch, 2018: 132).
[iii] For a discussion of pitfalls in the presentation of language data see Anthony 2018.
[iv] http://cass.lancs.ac.uk/.