

Accountability in corpus linguistics: A proposed framework and new tools for analysis

A new 'accountability' framework for use in corpus-based discourse analysis (and perhaps beyond)

Monika Bednarek (University of Sydney, E: monika.bednarek@sydney.edu.au), Martin Schweinberger (University of Queensland), Kelvin Lee (University of Sydney)

Introduction

Recent years have seen an increase in *meta-reflection*, i.e. reflection about corpus-based discourse analysis, covering concepts such as triangulation, objectivity/subjectivity, replication, transparency, reflexivity, and consistency (e.g. Marchi & Taylor 2009; Taylor & Marchi 2018; Baker & Egbert 2016)

Marchi and Taylor (2018: 12) argue for replacing the aim to achieve greater objectivity with the aim of increasing accountability (including consistency, transparency) and self-reflexivity. Bednarek et al (2024) propose a new *accountability* framework for use in corpus-based discourse analysis and beyond (<https://doi.org/10.1515/cllt-2023-0104>)

Accountability as a broad cover term (drawing on the everyday meaning) and as a multi-faceted phenomenon, covering various aspects of the research process.

"Being accountable means being transparent about research aspects, but also being able to justify them, being responsible for the decisions made or positions taken, and critically reflecting on them."

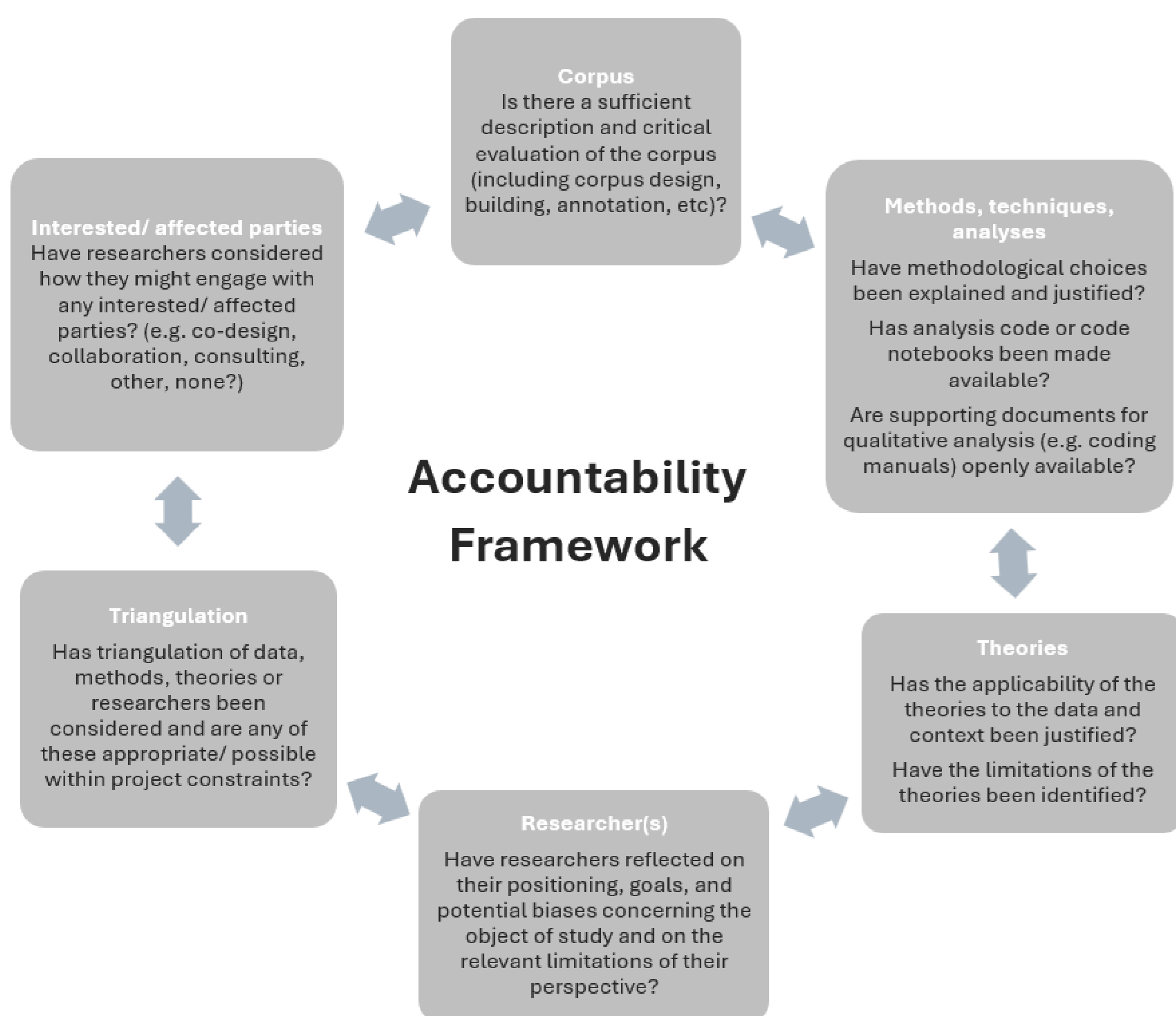
— Bednarek, Schweinberger, and Lee (2024)

Accountability framework

The Accountability framework can be understood as a set of guiding principles or as a reflection tool. But: the burden of implementation should not fall on individual researchers; rather, supporting structures, new initiatives, and cultural change are needed.

(See *Australian Text Analytics Platform*, www.atap.edu.au)

Find more information and the full paper using this QR-code.



Schematic representation of aspects associated with the accountability framework. Each component can be unpacked further with additional questions (in the full paper).

Australian Text Analytics Platform (ATAP)

ATAP is a new initiative which fosters cooperation among data and text analytics users and providers, and supports researchers from diverse academic backgrounds, including beginners, in adopting user-friendly code-based text analysis.



ATAP addresses one part of the accountability framework

(*Corpus; Methods, techniques, analyses*).

ATAP makes use of Jupyter notebooks to document, share, and perform code-based tasks relevant for corpus analyses.

Jupyter Notebooks

Jupyter notebooks are open-source web applications that allow users to create and share documents containing both text and executable code, supporting languages like Python, Julia, and R. As such, Jupyter notebooks are interactive documents that combine human-readable content with executable code.

These notebooks can be run either locally or on online platforms and can be converted to various file formats such as HTML, LaTeX, and PDF for easier sharing. A notebook is composed of three types of cells: code cells for running analyses, text cells for formatting text, and raw cells for configuration.

Example: ATAP Quotation Tool

Quotation is a crucial strategy in discourse representation, shaping perspectives and ideologies by illustrating how sources are portrayed and affect narrative framing.

QuotationTool

In this notebook, you will use the *QuotationTool* to extract quotes from a list of texts. In addition to extracting the quotes, the tool also provides information about who the speakers are, the location of the quotes (and the speakers) within the text, the identified named entities, etc., which can be useful for your text analysis.

Note: This code has been adapted (with permission) from the [GenderGapTracker GitHub page](#) and modified to run on a Jupyter Notebook. The quotation tool's accuracy rate is evaluated in [this article](#).

The notebook-based quotation tool automates the identification and classification of quoted sources, reporting expressions, and quotation types, facilitating the analysis of large datasets for biases and patterns.

The notebook enhances transparency by providing both code and explanations, allowing users to understand and adjust parameters by using data previews to spot issues.

The quotation extraction used in the notebook has been adapted (with permission) from code developed by Maite Taboada's Gender Gap Tracker team.

ATAP's Jupyter Notebooks

- Document Similarity Tool (identifies similar documents)
- Quotation Tool (extracts quotes and speakers)
- Semantic Tagger, Keywords Analysis, etc.
- NLP tools (e.g. topic modelling)
- LADAL Tools (see www.ladal.edu.au/tools.html): various notebook-based tools that perform basic tasks relevant for corpus analysis, e.g., concordancing, collocation analysis, keyword analysis, pos-tagging, etc.)

Full list at <https://www.atap.edu.au/resources/>

References

- 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.
- Taylor, C., & Marchi, A. (Eds.). (2018). *Corpus Approaches to Discourse: A Critical Review*. Routledge.
- Baker, P., & Egbert, J. (Eds.). (2016). *Triangulating Methodological Approaches in Corpus Linguistic Research*. Routledge.
- Bednarek, M., Schweinberger, M., & Lee, K. K. H. (2024). Corpus-based discourse analysis: From meta-reflection to accountability. *Corpus Linguistics and Linguistic Theory*. <https://doi.org/10.1515/cllt-2023-0104> (Published online April 16, 2024)



Acknowledgements

The Australian Text Analytics Platform (ATAP) receives funding from the Australian Research Data Commons (ARDC) via the HASS Research Data Commons and Indigenous Research Capability Program. The ARDC is funded by the National Collaborative Research Infrastructure Strategy (NCRIS).