Exploring academic stance on artificial intelligence through multidimensional keyword analysis

Margo Van Poucke Macquarie University, Australia

Artificial intelligence (AI) continues to spark debate in academia, especially regarding its implications for research and teaching. This study is motivated by the need to understand how researchers publicly articulate a standpoint on AI tools and systems. The concept of stance enables us to systematically trace how academic writers express certainty, affect and evaluation in public discourse (Biber & Finegan, 1988). Drawing on Systemic Functional Linguistics (SFL) (Halliday, 1993; Matthiessen, 2015) and the three metafunctions of language, the study explores how lexicogrammatical stance resources construe meaning and negotiate authority in academic news discourse, and contributes to register variation analysis (Moore, 2017).

Combining quantitative and qualitative approaches, the investigation examines stance-taking strategies in 172 online academic news articles, extracted from *The Conversation* (Biber & Egbert, 2016), and shows how multidimensional keyword analysis, using the Kullback-Leibler Divergence (KLD) and Gries' (2008, 2021) normalised dispersion measure (DPnorm), can uncover specific stance-taking strategies.

The target corpus (192,879 words), retrieved via Google Console, represents a hybrid, web-based sub-register blending informational and evaluative discourse. The reference corpus, extracted through NewsAPI (1,332,585 words), consists of online news articles prioritising information delivery and audience engagement. After extracting stance features with SpaCy and normalising keyword frequencies (Gabrielatos, 2018), the KLD measure is applied to identify key stance markers, complemented by Gries' DPnorm to assess their distribution throughout the texts. The study aims to address the following research objectives:

- To identify key lexicogrammatical stance markers in academic news discourse, categorising stance features into five groups (epistemic, attitudinal, modal, attributive/style/perspective and facilitation/effort/necessity/obligation) and corresponding subcategories (e.g. certainty, likelihood, attitude/perspective), based on Biber et al.'s (1999, 2006) framework, using human-supervised automated stance feature retrieval, syntactic parsing, KLD and tupleisation;
- 2. To explore the rhetorical and functional use of these stance markers in academic news articles by adding dispersion;
- 3. To demonstrate the feasibility and benefits (e.g. improved pattern detection, analysis of inter-registerial variation) of integrating relative entropy and dispersion in Corpus-Assisted Discourse Studies (CADS).

The analysis reveals that academic news writers frequently employ epistemic markers (*likely*, *suggest*) and modals (*can*, *could*) to express likelihood and caution, conveying impartiality and analytical rigour, especially with regard to AI's role in education. Attitudinal markers (*difficult*, *useful*) indicate evaluative stance, which appears to maintain a balance between optimism and critique. Dispersion analysis confirms that some markers are highly clustered, revealing salient rhetorical patterns. The markers are subsequently linked to the sub-register's communicative goals of information- and opinion-sharing and the authors' stance-taking strategies (Biber & Jones, 2005).

The combined use of quantitative analysis and qualitative interpretation demonstrates the complementarity of the approaches. Dispersion metrics prevent findings from being skewed by uneven distributions (Gries, 2021). The study further shows how computational methods can elucidate register-specific stance patterns, improving our understanding of how contextual factors may affect academic news discourse. The methodology provides a replicable framework for stance analysis in diverse online (sub-)registers and seeks to encourage interdisciplinary applications of corpus linguistics in SFL. Far from being at odds with in-depth discourse analysis, a mixed-methods CADS approach may synergistically deepen our insights into lexicogrammatical stance patterns by combining corpus-based metrics and thematic analysis grounded in SFL. In this case, it demonstrates how academic stance in public-facing discourse is strategically constructed through patterned lexicogrammatical choices, thus contributing to register analysis and an improved understanding of scholarly engagement with AI.

References

- Biber, D. (2006). Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), 97–116. <u>https://doi.org/10.1016/j.jeap.2006.05.001</u>
- <u>B</u>iber, D., & Egbert, J. (2016). Register variation on the searchable web: A multi-dimensional analysis. *Journal of English Linguistics, 44*(2), 95–137. https://doi.org/10.1177/0075424216628955
- Biber, D., & Finegan, E. (1988). Adverbial stance types in English. *Discourse processes*, *11*(1), 1- 34. https://doi.org/10.1080/01638538809544689
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education Limited.
- Biber, D., & Jones, J. K. (2005). Merging corpus linguistic and discourse analytic research goals: Discourse units in biology research articles. *Corpus Linguistics and Linguistic Theory*, 1(2), 151–182. https://doi.org/10.1515/cllt.2005.1.2.151
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor, & A. Marchi (Eds.), *Corpus approaches to discourse* (1st ed., pp. 225–258). Routledge. https://doi.org/10.4324/9781315179346-11
- Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International journal of corpus linguistics*, *13*(4), 403–437. https://doi.org/10.1075/ijcl.13.4.02gri
- Gries, S. T. (2021). Analyzing dispersion. In M. Paquot, & S. T. Gries (Eds.), *A practical handbook of corpus linguistics* (pp. 99–118). Springer International Publishing.

- Halliday, M. A. K. (1993). Quantitative studies and probabilities in grammar. In M. Hoey (Ed.), Data, description, discourse. Papers on the English language in honour of John McH. Sinclair. Harper Collins.
- Matthiessen, C. M. I. M. (2015). Halliday's conception of language as a probabilistic system.
 In J. J. Webster (Ed.), *The Bloomsbury companion to M.A.K. Halliday* (pp. 203–241).
 Bloomsbury.
- Moore, A. R. (2017). Register analysis in systemic functional linguistics. In T. Bartlett, & G. O'Grady (Eds.), *The Routledge handbook of systemic functional linguistics* (pp. 442-

461). Routledge.