

Multi-Dimensional analysis of literary style from a lexicogrammatical perspective

Carlos Kauffmann

São Paulo Catholic University
chkauffmann@corpuslg.org

In the area of Corpus Linguistics, studies that employed Multi-Dimensional (MD) analysis initially focused on the grammatical end of lexicogrammatical continuum (Berber Sardinha, 2020) to identify co-occurrent morpho-syntactic categories interpreted as dimensions of register variation (Biber, 1988). More recently, a new lexis-based MD framework has been introduced (Crossley & Louwse, 2007; Berber Sardinha, 2019; Zuppari, 2020), which uses lexical units as data. As a result, two major approaches to MD analysis exist: the grammatical MD analysis providing the grammatical dimensions, and the lexical MD analysis providing the lexical dimensions. As far as I am aware, just one MD study has attempted to integrate grammatical and lexical dimensions (Mayer, 2018), enabling a cline between lexis on one hand and structure on the other.

In this paper, I present the results of a study whose goal is to take a lexicogrammatical approach to MD analysis that brings together grammatical dimensions and lexical dimensions, thereby providing a holistic view of the use of lexicogrammatical resources in a corpus of literary texts. Most MD studies have characterized literary style from a grammatical perspective (Opas, 1992; Biber & Finegan, 1994; Egbert, 2012; Ali & Ahmad, 2016), and therefore a description of the full spectrum of lexicogrammatical resources used by literary authors is missing in the MD literature. In order to carry out the research, a corpus of novels and short stories by the most celebrated Brazilian fictional author, Machado de Assis (1839–1908), was collected. The corpus (called CLIMA) comprises 9 novels and 76 short stories, totaling 859,521 words. It was tagged for part of speech with the PALAVRAS parser, and lemmatized with TreeTagger for Portuguese. For reasons of space, readers are referred to Berber Sardinha and Veirano Pinto (2019) for a description of the methods used for both the grammatical and lexical MD analyses.

The grammatical MD analysis resulted in five dimensions: Abstract discourse vs. Orality, Narration, Hedging, Contextual vs. Dramatic discourse, and Thought presentation vs. Ornamentation. And the lexical MD analysis resulted in nine lexical dimensions: Sentimental expression, Romantic reference, Condition of man, Social representation, Urban setting, Patriarchal authority, Opposition, Metalanguage, and Uncertainty vs. Epistolary.

The integration of these grammatical and lexical dimensions was obtained through canonical correlation analysis (Afifi, May, & Clark, 2012; Mayer, 2018), which identified four significant canonical correlations, three of which were interpreted as aesthetic dimensions (Kauffmann & Berber Sardinha, 2021). Dimension 1, Formal Introspective Romanticism, emphasized hedging by the use of existence and modal verbs in tandem with romantic vocabulary (e.g., *heart, love*). Dimension 2, Sentimental Oralized Narrative, combined action verbs and conjunctions, which injected orality to the discourse, with lexis related to body parts and patriarchal authority (e.g., *sir, man*). Dimension 3, Dramatization, united thought presentation (mental verbs, first personal pronouns), orality (discourse markers), and metalinguistic vocabulary (e.g., *chapter, write*).

In conclusion, this study offered a lexicogrammatical-based interpretation about the unique style of Machado de Assis through Corpus Linguistics procedures that could be used as a blueprint for future stylistic studies about other authors, in different languages.

References

- Afifi, A., May, S., & Clark, V. A. (2012) *Practical Multivariate Analysis* (Fifth ed.). CRC Press.
- Ali, S. & Ahmad, S. (2016) Discourse style variation among the leading novelists of Pakistani fiction in English: A Multidimensional Analysis. *ELF Annual Research*, 18, 227-246.
- Berber Sardinha, T. (2019) Using Multi-Dimensional Analysis to detect representations of national cultures. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-Dimensional Analysis: Research methods and current issues* (pp. 231-258). Bloomsbury.
- Berber Sardinha, T. (2020) Lexicogrammar. In C. A. Chapelle (Ed.), *The Concise Encyclopedia of Applied Linguistics* (pp. 701-706). Wiley Blackwell.
- Berber Sardinha, T. & Veirano Pinto, M (Eds.), *Multi-Dimensional Analysis: Research methods and current issues* (pp. 231-258). Bloomsbury.
- Biber, D. (1988) *Variation across speech and language*. Cambridge University Press.
- Biber, D. & Finegan, E. (1994) Multi-dimensional analyses of authors' styles: some case studies from the 18th century. *Research in Humanities Computing*, 3, 3-17.
- Crossley, S. & Louw, M. M. (2007) Multi-dimensional register classification using bigrams. *International Journal of Corpus Linguistics*, 12(4), 453-478.
- Egbert, J. (2012) Style in nineteenth century fiction: A Multi-Dimensional analysis. *Scientific Study of Literature*, 2(2), 167-198.
- Kauffmann, C. & Berber Sardinha, T. (2021) Brazilian Portuguese literary style. In E. Friginal & J. A. Hardy (Eds.), *The Routledge handbook of corpus approaches to discourse analysis* (pp. 354-375). Routledge.
- Mayer, C. (2018) *O que e como escrevemos na Web: Um estudo multidimensional de variação de registro em língua inglesa [What and how we write on the Web: A multi-dimensional study of register variation in English]*. (Ph.D. dissertation.) Graduate Program in Applied Linguistics, São Paulo Catholic University, São Paulo.
- Opas, L. L. (1992) A multidimensional analysis of style in Samuel Beckett's prose works. *Research in Humanities Computing*, 4, 81-14.
- Zuppari, M. C. (2020) *Collocation dimensions in academic English*. (Ph.D. dissertation.) Graduate Program in Applied Linguistics, São Paulo Catholic University, São Paulo.