Multi-Dimensional Collocational Analysis of Discourses around COVID-19 Therapies

The goal of this study is to describe the discourses related to the endorsement or opposition of alternative COVID-19 treatments during the pandemic. This was achieved by compiling a corpus of academic articles that either advocated for or criticized the use of treatments like hydroxychloroquine and azithromycin, in accordance with recommendations from the WHO and other health organizations. The dataset was selected to represent both perspectives equally. Our methodology employed Lexical Multi-Dimensional Analysis (LMDA; Berber Sardinha & Fitzsimmons-Doolan, 2024), an offshoot of Multi-Dimensional Analysis (Biber, 1988, 1995; Berber Sardinha & Veirano Pinto, 2014, 2019). This approach focuses on lexical features (e.g., lemmas) and applies multivariate statistical techniques, such as Factor Analysis, to identify correlated lexical features across the texts. Specifically, we examined the keyword collocations within two subgroups: pro-alternative treatment (PAT) and against-alternative treatment (AAT). Keywords for each subgroup were identified by using the other subgroup as a reference corpus. We then determined the collocations for each keyword within a four-word window on either side, comparing these within both the PAT and AAT datasets. For each keyword, this process yielded two sets of collocates, one for PAT and one for AAT. To manage differences in sample size, we selected the 500 most significant collocates from each subgroup based on their logDice scores. This enabled a direct comparison of how keyword collocation changes with differing treatment perspectives. For example, the collocation shift for "hydroxychloroquine" includes its association with 'treatment' in PAT texts, highlighting potential benefits and supportive guidelines, whereas in AAT texts, it is linked to discussions of mixed results concerning recovery times and side effects. Another example of collocation shift refers to 'patients': in the PAT texts, its collocates suggest a focus on high-risk individuals needing urgent care or facing greater health risks; conversely, in the AAT texts, 'patients' are depicted as participants in research, emphasizing a scientific evaluation of treatments' effects, efficacy, and safety. Through LMDA, we identified the major dimensions in keyword collocate use across the corpus. Initial results include a dimension contrasting discourses: one promotes the extensive use of repurposed drugs, focusing on potential benefits and minimizing risks, despite uncertain evidence of their efficacy and safety; the other advocates for a cautious evaluation of outcomes like mortality and clinical improvement, highlighting the importance of transparency and ethical considerations in research. The paper will introduce, discuss, illustrate, and compare the dimensions based on treatment stance.

Key Words:

Multi-Dimensional Analysis, Register Variation, Artificial Intelligence

References:

Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, D. (1995). *Dimensions of Register Variation - A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.

Berber Sardinha, T., & Veirano Pinto, M. (Eds.). (2014). *Multi-Dimensional Analysis, 25 years on: A Tribute to Douglas Biber*. Amsterdam: John Benjamins.

Berber Sardinha, T., & Veirano Pinto, M. (Eds.). (2019). *Multi-Dimensional Analysis: Research Methods and Current Issues.* Abingdon: Bloomsbury.

Berber Sardinha, T., & Fitzsimmons-Doolan, S. (2024). *Lexical Multidimensional Analysis*. Cambridge: Cambridge University Press.