

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO**  
**SETOR DE PÓS-GRADUAÇÃO**

**Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem**

Disciplina: ATIVIDADE PROGRAMADA: Seminário de Pesquisa: Mapeamento do raciocínio de modelos de Inteligência Artificial com Análise Multidimensional

Linha(s) de Pesquisa: Linguagem e Tecnologia

Responsável: 006534 Tony Berber Sardinha

Créditos: 2

Semestre/Ano: 2025-1

Dia/Horário: Quinta-feira 16:00-19:00

Periodicidade:

Nível: ME/DO

ORCID: <https://orcid.org/0000-0001-8815-1521>

Research Gate: <https://www.researchgate.net/profile/Tony-Berber-Sardinha>

**Disciplina será oferecida em Português**

**Ementa:**

O objetivo do seminário é entender o raciocínio de Modelos Amplos de Linguagem (Large Language Models, LLMs), como GPT, Gemini e Llama. Por ‘raciocínio’ dos LLMs, entendemos a geração de conhecimento pelos modelos de IA. O problema de pesquisa relacionado à explicitação dos padrões de raciocínio da IA envolve dois fatores: a opacidade dos processos internos das redes neurais e a variabilidade natural das respostas dos LLMs. A metodologia empregada será a Análise Multidimensional (AMD), que permite identificar as dimensões de variação da língua em uso, seja ela natural ou sintética. Essas dimensões correspondem aos parâmetros subjacentes à linguagem. Dois tipos de AMD serão utilizados: a AMD Tradicional, que busca detectar os padrões comunicativo-funcionais da linguagem, e a AMD Lexical, que visa identificar os parâmetros discursivo-ideológicos embutidos na linguagem. A AMD é apropriada para a tarefa por ter como objetivo descrever a sistematicidade da variação linguística. A AMD parte do princípio de que o uso linguístico está fundamentado na variação natural entre textos, falantes, épocas e contextos, mas também pressupõe que, em meio a essa variabilidade, existem padrões recorrentes que revelam a sistematicidade do sistema linguístico.

**Cronograma de aulas:**

1. 6 de março de 2025
2. 13 de março de 2025
3. 20 de março de 2025
4. 27 de março de 2025
5. 3 de abril de 2025
6. 10 de abril de 2025
7. 24 de abril de 2025
8. 8 de maio de 2025
9. 15 de maio de 2025
10. 22 de maio de 2025
11. 29 de maio de 2025
12. 05 de junho de 2025

**Bibliografia:**

Berber Sardinha, T. (2022a). Corpus linguistics and historiography: Finding the major discourses in the first 50 years of TESOL Quarterly. *Journal of Research Design and Statistics in Linguistics and Communication Science*, 7(1), 69-90.

Berber Sardinha, T. (2022b). A text typology of social media. *Register Studies*, 4(2), 138-170.

Berber Sardinha, T. (2024a). AI-generated vs human-authored texts: A Multidimensional comparison. *Applied Corpus Linguistics*, 4(1), 100083. <https://doi.org/https://doi.org/10.1016/j.acorp.2023.100083>

Berber Sardinha, T. (2020a). A historical characterisation of American and Brazilian cultures based on lexical representations. *Corpora*, 15(2), 183-212.

Berber Sardinha, T. (2019). Using multi-dimensional analysis to detect representations of national culture. In T. Berber Sardinha & M. Veirano Pinto (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 231-258). London: Bloomsbury.

Berber Sardinha, T. (2021). Discourse of academia from a multi-dimensional perspective. In E. Friginal & J. Hardy (Eds.), *The Routledge Handbook of Corpus Approaches to Discourse Analysis* (pp. 298-318). Abingdon: Routledge.

Berber Sardinha, T. (2024). Exploring multimodal corpora in the classroom from a multidimensional perspective In P. Crosthwaite (Ed.), *Corpora for Language Learning: Bridging the Research-Practice Divide* (pp. 25-36). Abingdon: Routledge.

Berber Sardinha, T., & Fitzsimmons-Doolan, S. (in press). *Lexical Multidimensional Analysis: Identifying Discourses and Ideologies*. Cambridge: Cambridge University Press.