



**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: Linguística de corpus, estatística, multimodalidade

Linha(s) de Pesquisa: Linguagem e Tecnologia

Responsável: 006534 Tony Berber Sardinha

Créditos: 2

Semestre/Ano: 1º./2023

Dia/Horário: Quinta-feira das 16:00-19:00hs

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:**

Neste seminário, enfocaremos projetos do grupo de pesquisa GELC, lidando com questões relacionadas a design e análise de corpus, desenvolvimento e utilização de ferramentas de análise linguística e visual (computer vision), pacotes de análise estatística, procedimentos e testes estatísticos uni- e multivariados, processamento em nuvem (cloud computing), aprendizado de máquina (machine learning), análise multidimensional, variação de registro, anotação manual de dados, análise qualitativa, análise de discurso, padronização lexical, identificação de metáfora, mineração de dados, seleção de características (feature selection), etiquetagem e parsing, análise de sentimento, word embeddings, tipologia textual, tipologia discursiva, coleta em lote (scraping) de redes sociais, visão probabilística de linguagem, representatividade de corpus, variação, registro e padronização, bem como outros aspectos teóricos e empíricos necessários para a realização das pesquisas em andamento no grupo.

**Bibliografia:**

Berber Sardinha, T., & Veirano Pinto, M. (2021). A linguistic typology of American television. *International Journal of Corpus Linguistics*, 26(1), 127-160.

Berber Sardinha, T. (in press). A text typology of social media. *Register Studies*.

Biber, D. (1989). A typology of English texts. *Linguistics*, 27, 3-43.

Berber Sardinha, T., Kauffmann, C., & Acunzo, C. M. (2014). A multidimensional analysis of register variation in Brazilian Portuguese. *Corpora*, 9(2), 239-271.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

Berber Sardinha, T. (2021). Going multimodal in corpus linguistics: The case of social media. Invited talk presented at the International perspectives on corpus technology for language learning, University of Queensland.

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

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.