



Evento	Salão UFRGS 2022: SIC - XXXIV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2022
Local	Campus Centro - UFRGS
Título	Compilação de corpora de textos produzidos e avaliados nas edições 20152 a 20171 do Celpe-Bras
Autor	MILENA SOARES SILVA
Orientador	JULIANA ROQUELE SCHOFFEN

O exame Celpe-Bras (Certificado de Proficiência em Língua Portuguesa para Estrangeiros) tem como finalidade avaliar a capacidade dos examinandos de agir no mundo por meio da língua portuguesa e confere certificação em quatro níveis de proficiência: Intermediário, Intermediário Superior, Avançado e Avançado Superior. Estudos recentes envolvendo textos produzidos na Parte Escrita do exame (Divino, 2021) constataram que é possível verificar diferenças entre textos de diferentes níveis de proficiência utilizando métricas aferidas por programas de Linguística de Corpus, como já vem sendo feito em muitos exames (Glabasova *et al.*, 2019; Weigle *et al.*, 2013). A Linguística de Corpus pode ter papel significativo na análise de avaliações de larga escala, visto que examinam-se dados empíricos dos textos e, assim, torna-se possível descrever a performance linguística dos autores (Sardinha, 2004). O objetivo deste trabalho é descrever a elaboração de um corpus a partir da compilação de exemplos representativos dos textos produzidos e avaliados nas edições 2015/2, 2016/1, 2016/2 e 2017/1, contando com amostras de todos os níveis e de todas as tarefas dessas edições. O processo de compilação do corpus, ainda em desenvolvimento, conta com as etapas de digitação, desidentificação, revisão e edição. Os textos foram fornecidos pelo Inep em imagens digitalizadas. Durante a digitação, estão sendo omitidas quaisquer informações que possam identificar os autores dos textos. Posteriormente, cópias dos textos são editadas para padronizar a ortografia de acordo com a norma dicionarizada da língua portuguesa, com o intuito de evitar que as inadequações ortográficas influenciem a contagem de palavras (RÍO *et al.*, 2016). O corpus desenvolvido permitirá pesquisas inéditas sobre a avaliação realizada no Celpe-Bras, visto que possibilitará descrever os recursos léxico-gramaticais característicos de cada nível de proficiência, além de possibilitar outros estudos acerca do aprendizado de língua adicional (VOLODINA *et al.*, 2016).