



**REENCONTROS  
NOVOS ESPAÇOS  
OPORTUNIDADES**

**XXXIV SIC** Salão Iniciação Científica

**26 - 30  
SETEMBRO  
CAMPUS CENTRO**

<b>Evento</b>	Salão UFRGS 2022: SIC - XXXIV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2022
<b>Local</b>	Campus Centro - UFRGS
<b>Título</b>	Influência das inadequações ortográficas em análise de tarefa escrita do CELPE-Bras guiada por corpus
<b>Autor</b>	ISADORA DAHMER HANAUER
<b>Orientador</b>	JULIANA ROQUELE SCHOFFEN

Conforme os descritores de níveis do Certificado de Proficiência em Língua Portuguesa para Estrangeiros, textos do nível Avançado Superior - certificação máxima - podem apresentar pequenas inadequações que não comprometam a fluidez da leitura e a configuração da interlocução (BRASIL, 2020). Estudos realizados anteriormente (DIVINO, 2021; HANAUER, 2021; SILVEIRA, 2021) mostram que textos desse nível realizam pouca cópia explícita do material de insumo. Usando programas de análise de *corpus*, as autoras observaram que inadequações ortográficas estão presentes nas produções, podendo afetar os resultados das análises, impossibilitando o reconhecimento de palavras do texto de insumo. Neste trabalho, o *corpus* analisado foi composto por 240 textos escritos em resposta à Tarefa 2 da edição 2015-2 do exame, que integra compreensão oral e produção escrita. Os textos foram editados para padronizar a ortografia de acordo com a norma dicionarizada da língua portuguesa, para que palavras escritas com diferentes grafias fossem computadas como um só *type* pelo software utilizado, o *AntConc* (ANTHONY, 2019). Os mesmos textos foram anteriormente analisados sem alterações por Hanauer (2021), que verificou que os pacotes lexicais mais frequentes em comum com o insumo eram, em sua maioria, relativos ao título da tarefa. Em comparação, neste estudo, analisando os textos editados, observou-se que a lista de pacotes lexicais em comum com o insumo apresentou novas entradas e maior frequência nos pacotes mais usados, com expressões retiradas do conteúdo do áudio. Também foi calculado o *Type-Token Ratio*, índice que indica a densidade lexical dos textos (BIBER et al., 2002), sendo o resultado dos textos editados (6,65) menor que dos textos originais (7,68). É possível concluir que as inadequações ortográficas influenciam na métrica que possibilita a percepção da riqueza lexical, visto que os textos originais só têm o índice maior pois programas como o *AntConc* consideram como diferentes *types* palavras com grafias diferentes.