



Conectando vidas  
Construindo conhecimento



XXXIII SIC SALÃO INICIAÇÃO CIENTÍFICA

<b>Evento</b>	Salão UFRGS 2021: SIC - XXXIII SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
<b>Ano</b>	2021
<b>Local</b>	Virtual
<b>Título</b>	TAREFAS DE ESCRITA E COMPREENSÃO ORAL NO CELPE-BRAS: ANÁLISE DO AVANÇADO SUPERIOR BASEADA EM CORPUS
<b>Autor</b>	ISADORA DAHMER HANAUER
<b>Orientador</b>	JULIANA ROQUELE SCHOFFEN

## TAREFAS DE ESCRITA E COMPREENSÃO ORAL NO CELPE-BRAS: ANÁLISE DO AVANÇADO SUPERIOR BASEADA EM CORPUS

Autora: Isadora Dahmer Hanauer

Orientadora: Prof. Dra. Juliana Roquele Schoffen

Instituição: Universidade Federal do Rio Grande do Sul

No contexto do Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras), proficiência é a capacidade de engajamento e adequação às diferentes circunstâncias nas quais a língua portuguesa é utilizada no mundo. (BRASIL, 2020) O exame é composto de parte oral e parte escrita, sendo a segunda composta de quatro tarefas integradas de produção escrita e compreensão oral e escrita. Estudos recentemente realizados buscaram uma melhor caracterização dos níveis certificados pelo exame, tendo destaque Kunrath (2019), que analisou os recursos linguístico-discursivos empregados pelos examinandos, e Mendel (2019) abordando o conceito de autoria nos textos dos participantes. Desta maneira, o presente trabalho utilizou-se do software *Sketch Engine* (KILGARRIFF ET AL., 2004) para uma análise baseada em *corpus* de textos de participantes que atingiram o nível Avançado Superior nas tarefas 1 e 2 da edição 2015-2 do exame. Foram analisados 193 textos da tarefa 1 e 240 textos da tarefa 2, representando todos os textos que registraram nota máxima nestas tarefas. Os textos foram obtidos em forma de imagens digitalizadas, sendo digitados para compilar o corpus analisado. Nos *corpora*, as métricas identificadas foram as de pacotes lexicais (através do mecanismo n-grams) e expressões, em comparação com o texto de insumo da tarefa. Na tarefa 1, que tem insumo de vídeo, foi identificada uma grande quantidade de expressões em comum com o texto de insumo, bem como uma lista significativa de pacotes lexicais. Já na tarefa 2, expressões de duas a três palavras presentes também no insumo de áudio foram identificadas. Porém, a lista de pacotes lexicais foi menor. Considera-se como possível causa o uso de paráfrase, visto que observações qualitativas identificaram trechos similares aos do texto de insumo, que não apareceram na lista de pacotes lexicais. Futuramente, serão realizadas comparações entre mais tarefas e níveis de proficiência, e análises de outras métricas.