



XXXV SALÃO de INICIAÇÃO CIENTÍFICA

6 a 10 de novembro

Evento	Salão UFRGS 2023: SIC - XXXV SALÃO DE INICIAÇÃO CIENTÍFICA DA UFRGS
Ano	2023
Local	Campus Centro - UFRGS
Título	Desenvolvimento do protocolo de revisão e etiquetagem do Corpus de Produções Escritas do Celpe-Bras
Autor	ISABEL FRANCO LISBOA
Orientador	JULIANA ROQUELE SCHOFFEN

O certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras) é o exame brasileiro para a certificação de proficiência em Português como Língua Adicional (PLA). Este trabalho tem por objetivo apresentar o protocolo de digitação, revisão e etiquetagem desenvolvido pelo grupo Avalia. Esse protocolo irá permitir análises automatizadas do Corpus de Produções Escritas do Celpe-Bras (Schoffen et al, em preparação), compilado pelo grupo a partir de dados cedidos pelo Inep, que conta com cerca de 15.000 textos produzidos e avaliados em 4 edições do Celpe-Bras, com até 200 textos avaliados em cada nota (0-5) por cada tarefa (4 por edição), estimando-se totalizar cerca de 3 milhões de palavras. O processo de digitação e revisão respeita a escrita original do texto e exclui qualquer marca de identificação. A etiquetagem é feita manualmente através de VBA (Visual Basic for Applications) no Microsoft Word, seguindo as sugestões de Hardie (2014) para uso de XML simplificada. O protocolo estabelece regras para a marcação de diferenças ortográficas visando a sua padronização, de forma a permitir seu reconhecimento por ferramentas automatizadas e posterior análises de riqueza lexical que não sofram distorções por conta de diferentes grafias da mesma palavra (Hanauer, 2022). Além disso, considerando o construto teórico do Celpe-Bras, que leva em conta a adequação ao contexto comunicativo proposto, o protocolo inclui marcações de aspectos representativos da relação de interlocução configurada nos textos dentro dos gêneros solicitados nas tarefas, tais como sinalização de características estruturais (por exemplo, abertura, cabeçalho, endereçamento e saudações). Além de contribuir também para a discussão sobre etiquetagem e análise de textos produzidos em contextos semelhantes, a compilação e subsequentes análises do corpus de produções escritas do Celpe-Bras visam refinar a descrição dos níveis de proficiência em língua portuguesa, buscando ampliar as evidências de validade do Celpe-Bras.