

## **Corpus TecEM: uma plataforma digital de textos produzidos por alunos de cursos técnicos integrados ao ensino médio**

Bruno Corrêa de Almeida<sup>1</sup>, Pedro de Andrade Santos<sup>1</sup>, Maitê Moraes Gil<sup>\*</sup>  
\*Orientadora

<sup>1</sup>Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) -  
*Campus Osório. Osório, RS*

A Linguística de Corpus (Ldc) é uma abordagem empírica para o estudo da língua que se fortaleceu nas últimas décadas. Ela investiga o que é teoricamente possível na língua e tem como foco a investigação do uso e da maneira como os usuários utilizamos recursos de linguagem disponíveis, através da observação de material autêntico. Na Ldc, um corpus é uma coleção de textos produzidos naturalmente na língua (em contraposição a textos induzidos e à língua da máquina), armazenados em formato eletrônico e com o intuito de serem alvo de investigações linguísticas. O projeto de pesquisa Corpus TecEM visa a construir uma plataforma digital cujo conteúdo é um conjunto de textos produzidos por alunos de cursos técnicos integrados ao ensino médio. Sendo o corpus uma ferramenta desenvolvida para professores e pesquisadores, buscamos atender algumas das suas necessidades e oferecer ferramentas como métodos de filtragem dos textos de acordo com seus interesses para ações de planejamento de aulas e investigações linguísticas. Justifica-se o interesse em produções de alunos de cursos técnicos integrados ao Ensino Médio com o fato de este ser um espaço recente de estudos, o qual se fortaleceu com a expansão dos Institutos Federais. A metodologia deste projeto se baseia principalmente nos princípios da Linguística de Corpus, e se guia pelas seguintes etapas: (i) levantamento bibliográfico (ii) desenvolvimento da ferramenta; (iii) contato com professores de Língua Portuguesa de Institutos Federais, para apresentação da proposta da pesquisa e convite aos interessados em contribuir na coleta de textos; (iv) coleta de textos escritos por alunos de cursos técnicos integrados ao Ensino Médio durante suas aulas de Língua Portuguesa ao longo do ano; (v) compilação dos textos a partir de critérios estabelecidos nos passos iniciais deste projeto; (vi) armazenamento do Corpus TecEM em um banco de dados online; e (vii) disponibilização da ferramenta. As três primeiras etapas já foram desenvolvidas, e as etapas (iv), (v) e (vi) estão em andamento. O site atualmente conta com a página inicial, que contém informações sobre o corpus, o posicionamento provisório do concordanciador e uma área de login para os administradores. Ao final do projeto, espera-se ter disponibilizado à comunidade acadêmica uma base de dados rica e criteriosa para futuras pesquisas, contribuindo tanto para o desenvolvimento teórico quanto para a proposição de novas práticas de ensino de Língua Portuguesa.

**Palavras-chave:** Corpus. Plataforma Digital. Análise Linguística.