

Utilizando PLN para analisar a utilização de resultados de avaliação por parte dos professores

João Augusto Ferreira Freire – CAEd/UFJF joao.freire@fundacaocaed.org.br

Luiz Vicente Fonseca Ribeiro – CAEd/UFJF tuko@fundacaocaed.org.br

Fillipe Guedes Soares – CAEd/UFJF fillipe.soares@fundacaocaed.org.br

Virgínia Rodrigues Strack – CAEd/UFJF virginia.strack@fundacaocaed.org.br

Natasha Ferrari Lopes – CAEd/UFJF natasha.lopes@fundacaocaed.org.br

Resumo

Este estudo investiga o nível de utilização e interpretação dos resultados de avaliações externas por professores, com base nas respostas ao item: "Como você interpreta e utiliza os resultados das avaliações dos estudantes?", extraído de um questionário contextual do CAEd/UFJF. Utilizando técnicas de Processamento de Linguagem Natural (PLN) e métodos computacionais, a pesquisa adotou uma abordagem exploratória, categorizando as respostas por meio de um score de uso e interpretação, calculado a partir da similaridade de Jaccard com 25 "frases-ouro" ideais. Além disso, a análise de tópicos latentes (LDA) mapeou os temas predominantes. Os resultados mostraram que 72,8% dos professores se concentram nos dois níveis mais baixos do score, enquanto apenas 3,5% atingiram o nível mais alto. Foram identificados dois tópicos principais: reflexão e ajuste pedagógico (Tópico 1) e práticas cotidianas de ensino e avaliação (Tópico 2). Conclui-se que há uma utilização limitada dos resultados para melhorias pedagógicas, indicando a necessidade de capacitação docente em interpretação de dados.

Introdução

A avaliação docente tem se consolidado como uma ferramenta estratégica para o aprimoramento da qualidade educacional, fornecendo subsídios para reflexão sobre práticas pedagógicas e desempenho discente. No entanto, a interpretação e utilização efetiva desses resultados pelos professores ainda representam um desafio, demandando abordagens inovadoras para análise qualitativa de seus relatos. Nesse contexto, o Processamento de Linguagem Natural (PLN) emerge como uma ferramenta promissora, permitindo a exploração sistemática de grandes volumes de dados textuais por meio de técnicas computacionais. Este estudo utiliza métodos de PLN, como similaridade textual e modelagem de tópicos, para analisar como os professores interpretam e aplicam os resultados de avaliações em sua prática, oferecendo assim novas perspectivas para compreender e potencializar esse processo no âmbito educacional.

Objetivos

O objetivo desta análise é avaliar o grau de utilização e interpretação dos resultados de avaliações externas por professores respondentes e mapear possíveis formas de utilização desses resultados, com em respostas ao item “Como você interpreta e utiliza o resultado das avaliações realizadas pelo estudante?”, presente no questionário contextual do Centro de Políticas Públicas e Avaliação da Educação da Universidade Federal de Juiz de Fora (CAEd/UFJF).

Este estudo apresenta um viés exploratório, sem tentar associar os indicadores construídos aos resultados de proficiência dos estudantes. A abordagem proposta é mais voltada para a identificação e o mapeamento de padrões específicos dentro das respostas dos professores, buscando construir um ponto de partida para a análise e interpretação de dados advindos de itens de resposta construídas aplicados a profissionais da educação. Almejou-se categorizar as respostas dos professores, por meio de um *score* de utilização e interpretação para cada resposta fornecida, com base em técnicas de Processamento de Linguagem Natural (PLN) e métodos computacionais. Ademais, o mapeamento das formas de utilização foi feito por meio de análise de tópicos latentes, utilizando o modelo *Latent Dirichlet Allocation* – LDA (Blei et. al, 2003).

Metodologia

No presente estudo, trabalharemos com duas frentes de análise do campo do PLN: *i*) a criação de um *score* de utilização e interpretação dos resultados de avaliações por meio do índice de semelhança de Jaccard; e *ii*) o mapeamento do vocabulário dos professores por meio da análise de tópicos latentes via LDA.

Com relação à primeira frente, resultados preliminares indicaram que a fase do pré-processamento textual acabava por diluir a nuance das respostas. Portanto, decidimos não alterar nem simplificar as respostas dos professores (ou seja, não aplicar o pré-processamento textual), pois constatamos que esses ajustes, nesse caso, acabavam prejudicando a identificação de palavras importantes por parte do modelo. No entanto, utilizamos um modelo de embeddings pré-treinado em português (*pt-core-news-lg*) como método de vetorização das respostas (Garcia et. al, 2024).

Para a criação do *score* de uso e interpretação dos resultados de avaliações, criamos um conjunto de 25 frases consideradas ideais, ou “frases-ouro”, que representam exemplos de boas práticas e reflexões que mostram uma utilização significativa dos resultados das avaliações para ajustar estratégias de ensino. A ideia é cobrir diversos estilos de escrita, práticas pedagógicas e formas diferentes de falar sobre o mesmo assunto.

Por fim, a semelhança textual foi calculada pelo método de similaridade de *Jaccard* (HUANG,2008) entre os vetores ajustados ao modelo de *embeddings*, para cada professor *i* e cada frase-ouro *k*. Sendo A_i a resposta individual de cada professor *i*, e B_k cada uma das *k* frases-ouro, calculamos o índice de Jaccard:

$$J(A_i, B_k) = \frac{A_i \cap B_k}{A_i \cup B_k} \quad (1)$$

Em outras palavras, cada resposta real dos professores foi comparada a cada uma das frases-ouro para medir o grau de semelhança, gerando 25 *scores* de semelhança. Então, fizemos a média dos 25 *scores* e normalizamos os resultados entre 0 e 1. A partir dessa escala, dividimos os *scores* em quatro níveis em ordem crescente, com cortes em 0,25, 0,50 e 0,75. A distribuição das respostas dos professores nos níveis é representada na tabela a seguir.

Partindo para a segunda etapa da análise, exploramos os temas predominantes em suas respostas, com ênfase nas práticas pedagógicas, no uso de avaliações externas e nas estratégias de ensino. Para essa análise, aplicamos o modelo *Latent Dirichlet Allocation (LDA)*, um método que busca identificar tópicos. Em termos simples, o *LDA* identifica padrões ao observar quais palavras costumam aparecer juntas, formando assim tópicos latentes, ou seja, conjuntos de termos frequentemente associados. A partir dos tópicos identificados e palavras mais frequentes por tópico, a interpretação seguiu uma lógica qualitativa, respeitando o contexto da avaliação educacional.

Resultados

Abaixo, segue a distribuição dos professores de acordo com o nível do índice criado.

Tabela 1 – Distribuição das respostas dos professores no índice de uso e interpretação dos resultados das avaliações

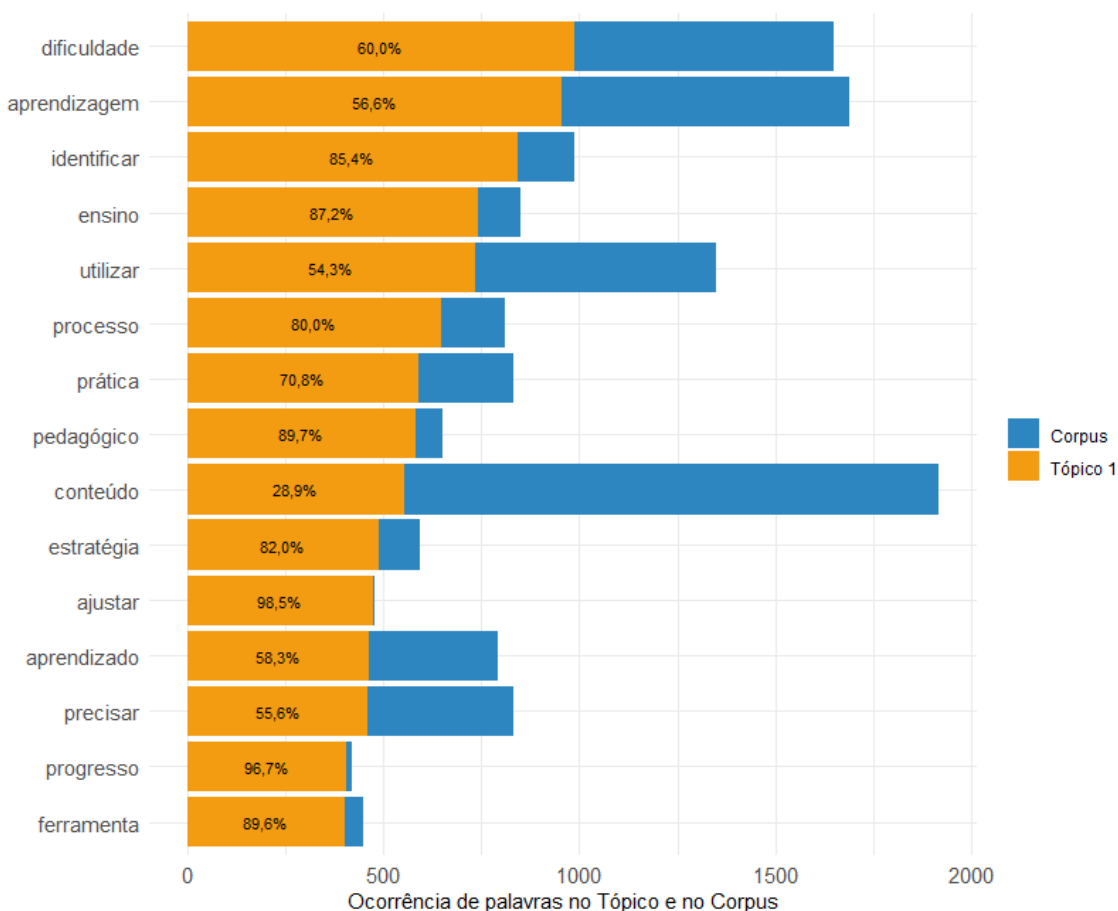
Nível do Índice de Utilização e Interpretação	Nº de Respostas	Percentual de Respostas
1	2652	29,5%
2	3898	43,3%
3	2135	23,7%
4	318	3,5%

Percebe-se que 72,8% dos professores se encontram nos dois primeiros níveis e apenas uma pequena parcela dos professores se encontra no nível mais elevado do índice, representando menos de 4% da população. Partindo para o mapeamento do vocabulário dos professores, a escolha do número de tópicos para o modelo foi auxiliada por métricas de

coerência e pela visualização do espaço intertópico gerado pela ferramenta *pyLDAvis*, sendo o modelo de 2 tópicos considerado o mais adequado.

A seguir, apresentamos um gráfico com a frequência relativa dos termos para cada tópico com relação ao *corpus*. Nos gráficos, o eixo “x” representa o número de vezes que a palavra aparece. A barra azul, completa, representa o número de ocorrência da palavra no *corpus* completo (ou seja, sem diferenciar pelos tópicos). A barra menor, por sua vez, representa o número de vezes que a palavra aparece para o respectivo tópico. Por fim, a porcentagem dentro da barra menor é a proporção de ocorrências da respectiva palavra dentro do tópico específico, com relação ao total. Ou seja, é a proporção de ocorrências daquela palavra, no tópico específico, com relação ao *corpus*. Abaixo, apresentamos o gráfico de barras empilhadas com a frequência relativa de palavras por tópico para o 1º Tópico analisado.

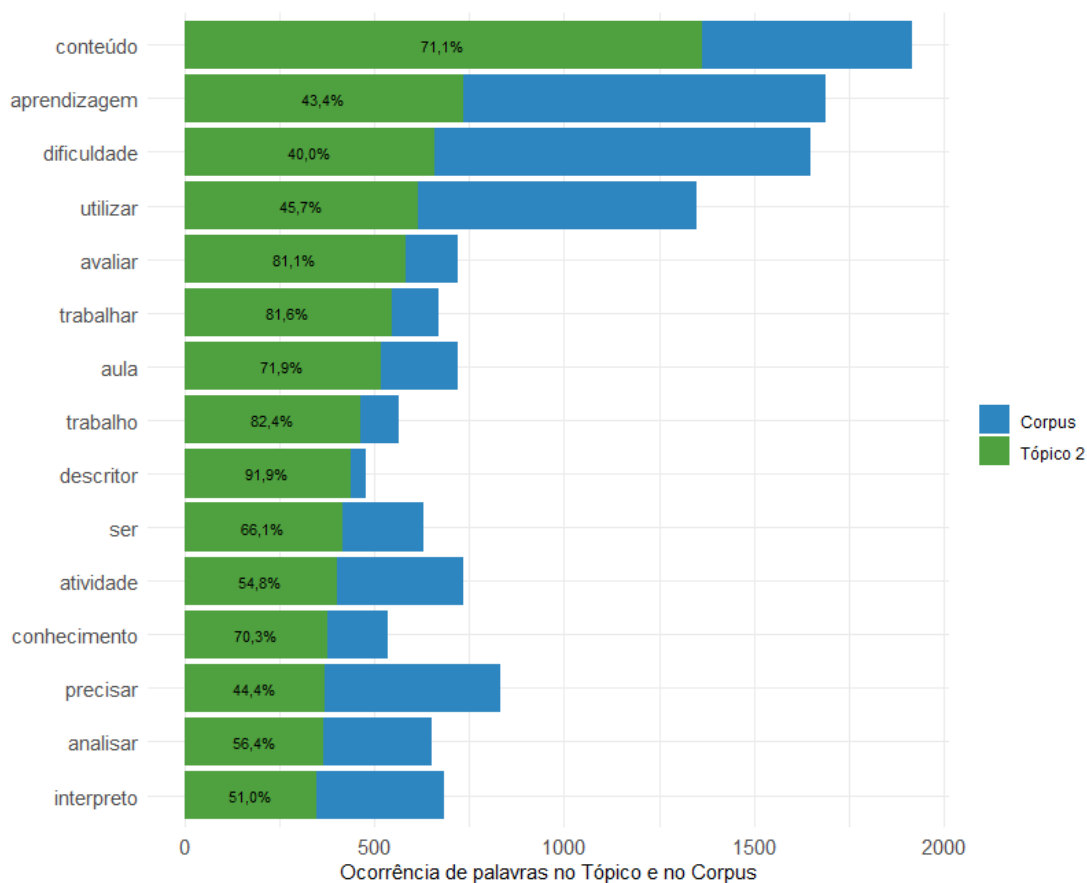
Gráfico 1 – Frequência de termos para o Tópico 1



O Tópico 1 apresenta um conjunto de palavras que indicam alguma ênfase na reflexão pedagógica, no processo de ensino-aprendizagem, e na utilização de avaliações como ferramentas para ajustar práticas educacionais. As palavras mais frequentes (com frequência relativa maior ou igual a 70%) são: 'ajustar' (98,5%), 'progresso' (96,7%), 'pedagógico' (89,7%), 'ferramenta' (89,6%), 'ensino' (87,2%), 'identificar' (85,4%), 'processo' (80%), 'prática' (70,8%)

e 'estratégia' (82%). Essas palavras sugerem que os professores estão focados em como utilizar os resultados das avaliações para melhorar suas práticas pedagógicas, identificar dificuldades dos alunos e acompanhar seu progresso.

Gráfico 2 – Frequência de termos para o Tópico 2



No Tópico 2, as palavras mais empregadas, com frequência relativa igual ou superior a 70%, são 'trabalho' (82,4%), 'avaliar' (81,1%) e 'aula' (71,9%). Essas palavras sinalizam que os professores se preocupam com a rotina diária, com a utilização de avaliações e o planejamento de aulas. Isso sugere que as práticas pedagógicas incluem tanto a transmissão de conhecimento quanto a avaliação baseada em descritores claros.

Discussão

No geral, os resultados mostram um baixo nível de utilização e interpretação dos resultados de avaliação educacional. O Tópico 1 suscita uma preocupação com a identificação de dificuldades e o acompanhamento do progresso, o que indica uma utilização das avaliações para monitorar o aprendizado e ajustar práticas. A interpretação desses resultados parece estar voltada para a reflexão sobre como os alunos estão se desenvolvendo e quais adaptações são necessárias no ensino.

No geral, o tópico revela que há reflexão sobre as próprias práticas pedagógicas. As palavras 'ajustar', 'estratégia' e 'prática' pode indicar que os professores estão engajados em um processo contínuo de autoanálise e melhoria, utilizando as avaliações como fonte de inspiração para tornar seu trabalho mais efetivo.

Acerca do Tópico 2, a palavra 'trabalho' pode estar relacionada à preparação e execução de atividades pedagógicas, enquanto 'avaliar' aventa uma preocupação com a avaliação dos alunos. 'Aula' acompanha a ideia de que as práticas pedagógicas estão centradas nas sessões de ensino regulares. Outras palavras importantes incluem 'conteúdo' (71,1%) e 'descriptor' (91,9%), podendo indicar que os professores estão preocupados com o material didático e o uso de critérios específicos para avaliar os alunos, respectivamente.

Por fim, o Tópico 2 indica que o foco pode estar nas atividades diárias de ensino e na avaliação baseada em critérios específicos. Também sugere o uso de conteúdo relevante e a identificação de dificuldades dos alunos, mas talvez uma reflexão menos aprofundada sobre como adaptar as práticas pedagógicas com base nos resultados das avaliações.

Conclusão

Os resultados deste estudo indicam que a maior parte dos professores (72,8%) apresenta níveis moderados de utilização e interpretação dos resultados de avaliações, enquanto uma parcela menor (3,5%) demonstra um engajamento mais avançado. A análise de tópicos revelou duas tendências principais: uma relacionada à reflexão sobre práticas pedagógicas (Tópico 1) e outra voltada para o planejamento de aulas e avaliações (Tópico 2), sugerindo diferentes abordagens na incorporação desses dados ao trabalho docente.

Em relação aos aspectos metodológicos, observou-se que o pré-processamento textual exigiu ajustes para preservar as nuances das respostas, o que pode ser considerado em pesquisas futuras. A definição do número de tópicos no modelo LDA também apresentou desafios, exigindo equilíbrio entre precisão técnica e clareza interpretativa. Esses achados reforçam a importância de aprimorar tanto as ferramentas de análise quanto as estratégias de apoio aos professores, de modo a potencializar o uso pedagógico dos resultados de avaliações.

Referências

BLEI, David M. et al. Latent Dirichlet Allocation. **Journal of Machine Learning Research**, v. 3, p. 993-1022, 2003. Disponível em: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf> Acesso em: 21 maio 2025.

GARCIA, Klaifer. et al. Breaking news: Unveiling a new dataset for Portuguese news classification and comparative analysis of approaches. **PLoS ONE**, v. 19, n. 1, e0296929, 2024.

Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0296929>
Acesso em: 21 maio 2025.

HUANG, Anna et al. Similarity measures for text document clustering. In: **Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand**. 2008. p. 9-56. Acesso em: 23 maio 2025.