



ANAIS

Simpósio Internacional de Língua, Literatura e
Interculturalidade (SIELLI)
e Encontro de Letras

www.sielli.ueg.br

POSLLI
PROGRAMA DE PÓS-GRADUAÇÃO
EM LÍNGUA, LITERATURA E INTERCULTURALIDADE

Câmpus
Cora Coralina

Universidade
Estadual de Goiás

09 A 13 DE NOVEMBRO DE 2020

EFEITO DA QUALIDADE LEXICAL NA AVALIAÇÃO DE REDAÇÕES EM LÍNGUA PORTUGUESA

EFFECT OF LEXICAL QUALITY ON THE ASSESSMENT OF ESSAYS IN PORTUGUESE

Jeane Cardoso Costa¹
Eduardo Batista da Silva²

Resumo:

A presente pesquisa investiga o uso de vocabulário mais formal em redações em língua portuguesa, com base em cálculos estatístico-lexicais. A fundamentação teórica foi orientada pela Linguística de Córpus (BERBER SARDINHA, 2012; BIBER; REPPEN, 2015). Nosso objetivo é verificar a relação entre o vocabulário mais formal e a nota dos corretores. Utilizamos o *software* Wordsmith Tools para o processamento de 10 redações. Participaram da pesquisa 10 professores de produção textual. Os resultados mostram que existe uma correlação positiva entre a quantidade de palavras mais formais no texto e a nota dos corretores, ou seja, quanto mais palavras formais, maior tende a ser a nota atribuída. Esperamos trazer contribuições para os estudos lexicais e estratégias de ensino que envolvem produção textual.

Palavras-chave: Vocabulário. Redações. Avaliação. Linguística de Córpus.

Abstract:

The present research investigates the use of formal vocabulary in essays in Portuguese, based on lexical statistical calculus. The theoretical framework was based on Corpus Linguistics (BERBER SARDINHA, 2012; BIBER; REPPEN, 2015). We aimed at checking the relationship between formal vocabulary and reviewers' grades. We used the software Wordsmith Tools so as to process 10 essays. Ten writing teachers participated in the research. Our findings show that there is a positive correlation between the number of formal words in the text and the grade assigned by the reviewers, that is, the more formal words, the more grades tend to be higher. We expect to contribute to lexical studies and teaching strategies involving writing.

Keywords: Vocabulary. Essays. Assessment. Corpus Linguistics.

Introdução

Haja vista as dificuldades e desafios ligados ao domínio de vocabulário em língua portuguesa, em especial à competência e habilidade escrita, um problema recorrente que merece discussão é a variabilidade desse vocabulário. Como recorte de pesquisa, optamos por explorar a qualidade de vocabulário de redações escritas em língua portuguesa. O presente trabalho

¹ Mestranda do Programa de Pós-Graduação *Stricto Sensu* "Língua, Literatura e Interculturalidade", da UEG/Câmpus Cora Coralina. E-mail: jeanecardosocosta@uol.com.br.

² Doutor em Estudos Linguísticos pela UNESP. Professor permanente do Programa de Pós-Graduação *Stricto Sensu* "Língua, Literatura e Interculturalidade", da UEG/Câmpus Cora Coralina. E-mail: eduardo.silva@ueg.br.



09 A 13 DE NOVEMBRO DE 2020

diferencia-se de trabalhos já realizados na área de avaliação de redações em língua portuguesa por duas razões principais: recorre à ferramentas de processamento lexical e baseia-se em cálculos estatístico-lexicais para possibilitar uma análise quantiquantitativa.

O tema discutido aqui acerca da avaliação do vocabulário contido em redações de alunos do Ensino Médio mostra-se relevante, pois busca trazer contribuições aos estudos lexicais e da Linguística de Córpus a partir da percepção de um fenômeno importante e atual.

Partindo ainda do pressuposto de que nas correções de redações de Ensino Médio, os quesitos mais considerados são a adequação ao tema e gênero, a capacidade de argumentação, a coesão e coerência estabelecidas no texto cuja preocupação maior é o modo como as ideias estão sendo estruturadas e como os conectivos estão sendo usados, surge a necessidade de pensarmos de modo aprofundado acerca de como a perspectiva lexical tem sido considerada nessas correções. Em outras palavras, como tem sido observado o nível do vocabulário do aluno ao empregar a língua escrita nesse contexto.

É importante ressaltar, contudo, que alguns esforços estão sendo feitos no sentido de contribuir com essa ampliação lexical e com o bom uso de vocabulário. Isso tem sido feito por meio de ferramentas interativas de divulgação que podem ser acessadas por grande parte dos alunos. O Instituto Nacional de Estudos e Pesquisas (INEP), por exemplo, em sua página *online* oficial tem compartilhado de ideias acerca das descobertas vocabulares por parte dos alunos, chamando atenção à necessidade de ampliação do insumo lexical para a boa escrita e interpretação de textos.

Nossa pesquisa não visa a caracterização de erros, tampouco destacar deficiências, mas, sim, angariar esforços e apresentar aos profissionais que lidam com correção de redações possibilidades para eventualmente agilizar o trabalho, pautadas no processamento linguístico-computacional – possibilitando a otimização do trabalho docente.

Frente ao exposto, ao tratarmos de análises qualiquantitativas, a investigação procura trazer contribuições para os professores já atuantes e para professores em formação inicial. Nosso objetivo é verificar a relação entre o vocabulário mais formal e a nota dos corretores.

Outrossim, uma vez que não são muitos os trabalhos acerca desse tema, essa temática tende a ser útil a pesquisadores, estudantes e professores que desejem compreender ou aprofundar conhecimentos nessas áreas.

Fundamentação Teórica

O aporte teórico de nossa investigação recorre principalmente aos estudos da Linguística de Córpus. Expomos, na sequência, uma breve apresentação dessa linha de pesquisa.

Linguística de Córpus

Iniciamos nossa argumentação apresentando uma definição de córpus: uma coletânea de textos disponíveis em formato eletrônico. Por córpus, entendemos como sendo um conjunto de textos que podem ser escritos, falados ou multimodais. Esses textos são coletados de forma criteriosa e mantidos em formato eletrônico, constituindo o alvo da pesquisa linguística. Os



09 A 13 DE NOVEMBRO DE 2020

dados constituem um *córpus* sistematizado, atendendo a alguns critérios amplos e profundos, de tal forma que representem o uso linguístico, conforme a conceituação de Sinclair (1991):

Um *corpus* é uma coletânea de textos autênticos, escolhidos para caracterizar um estado ou variedade de uma língua. Na Linguística Computacional moderna, um *corpus* tipicamente contém muitos milhões de palavras; isso se deve à criatividade da língua natural leva à uma imensa variedade de expressão. Torna-se difícil isolar os padrões recorrentes que são os indicadores da estrutura lexical de uma língua³. (SINCLAIR, 1991, p. 171, tradução nossa).

Córpus são usados para conhecimentos sobre a língua que podem complementar ou suplantam informações oriundas de fontes de referência e de introspecção (LEECH, 1991, 1992). Os *córpus* escritos ainda dominam a produção na área e têm encontrado uma aplicabilidade crescente não somente nos estudos linguísticos canônicos, mas no desenvolvimento de tecnologias da fala, do reconhecimento e da síntese.

A afirmação de que todo *córpus* traz questões novas ou questões que não imaginávamos encontrar é amplamente conhecida, ainda que nenhum *córpus* nos responda tudo. Desse modo, tanto as observações como os experimentos e hipóteses inicialmente formuladas no âmbito da investigação, nos orientam à uma revisão à luz das comprovações e dos resultados.

Assim sendo, sistematizar dados e observações é um ponto crucial, talvez mais importante do que a simples aplicação e contraste de teorias. Descobrir e identificar padrões a partir da observação são os problemas fundamentais, levando à uma compreensão do uso linguístico – uma vez que a língua constitui-se como um sistema probabilístico de combinatórias no qual as associações mantidas com outras unidades revelam estruturas recorrentes.

No contexto de ensino, o trabalho com *córpus* pode fornecer subsídios para a sala de aula. Nesse sentido, Berber Sardinha (2010) traz considerações importantes acerca de como preparar materiais de ensino de línguas com *córpus*. Para que as aulas aconteçam, há uma condição essencial: a preparação de materiais de ensino. Contudo, é necessário lembrar que para o ensino seja concretizado na perspectiva defendida aqui, o professor não pode ser dependente de si mesmo, de sua eloquência ou de um quadro negro, é necessário ter dados concretos oriundos de um *córpus*. Berber Sardinha (2010) acrescenta que existe uma metáfora de que o professor é um artista, além dessa existem outras subjacentes ao trabalho docente. Se cada aula é única e possui muitas horas, estaria essa arte sendo feita com alta qualidade? A prática docente deve estar alicerçada em artefatos prontos, rápidos e reutilizáveis.

³ A *corpus* is a collection of naturally-occurring language text, chosen to characterize a state or variety of a language. In modern computational linguistics, a *corpus* typically contains many millions of words; this is because it is recognized that the creativity of natural language leads to such immense variety of expression. That is difficulty to isolate the recurrent patterns that are the clues to the lexical structure of the language



09 A 13 DE NOVEMBRO DE 2020

Além disso, pelo que sabemos, cada aula é única e irreproduzível; então, nesse sentido o trabalho do professor é essencialmente uma arte mesmo. Porém, a realidade também nos mostra que o dia a dia da vida do professor é feito de muitas horas em sala de aula, diante do que é quase impossível ser artista continuamente e produzir arte de boa qualidade. As duas metáforas (do profissional e do artista) estão presentes, se complementam e entram em conflito, muitas vezes. (BERBER SARDINHA, 2010, p. 294).

Viana (2011) afirma que como um campo de conhecimento dedicado à investigação da linguagem, a Linguística não se distancia da prática de professores de línguas. Essa é uma constatação lógica (científica e pedagógica), pois voltam-se ao mesmo objeto de ensino e de análise, embora existam ainda profissionais que vêem a Linguística meramente como uma matéria cursada no currículo universitário. Todavia, estudar a linguagem pressupõe a prática docente. Para compreender como os alunos expressam as suas ideias, é preciso entender o funcionamento dela.

Recorrendo a Biber e Reppen (2015) para conceituar a Linguística de Córpus, entendemos que trata-se de uma abordagem de pesquisa que facilita as investigações empíricas relacionadas à variação e ao uso da língua. No que se refere à aproximação com o processamento de redações, vale salientar que tal abordagem permite um estudo empírico com implicações para o ensino.

A Linguística de Córpus se ocupa do córpus, sua análise e compilação. O termo Linguística de Córpus, segundo Sarmento (2010), pode ser compreendido como um estudo da linguagem que baseia-se em exemplos do cotidiano real.

A Linguística de Córpus não pode ser vista como a sintaxe, a semântica e a pragmática. É um ramo da Linguística que concentra-se na descrição e explicação de algum aspecto linguístico em uso, ela é uma metodologia que pode ser aplicada a uma gama de estudos linguísticos, até mesmo ao ensino de línguas. Ou seja, é uma das várias maneiras de se fazer Linguística.

Biber, Conrad e Reppen (1998, p. 4) trazem algumas características cruciais na Linguística baseada em córpus: ela é empírica, pois analisa padrões reais de uso em textos naturais; utiliza a coletânea de textos para análise; faz uso de computadores, usando também técnicas automáticas e interativas; é dependente de técnicas analíticas, quantitativas e qualitativas para gerar conhecimento empírico.

Para Novodvorski e Finatto (2014), é importante dizer que a Linguística de Córpus se coloca como uma nova perspectiva para a Linguística, mas não como um novo tipo de Linguística. Mostra-se àqueles que se aproximam dela como uma metodologia ou como uma abordagem teórica diferenciada dos estudos da linguagem. Se alguém quiser se aproximar da Linguística de Córpus apenas por ter um interesse em seu instrumental ou em seus procedimentos, não será nada cobrado em termo de filiação teórica ou epistemológica. Ainda que de modo insistente, é importante ressaltar que a Linguística de Córpus é também um modo de compreender a língua, um modo próprio de defini-la como objeto de estudo.

Moore (2003) elenca algumas possibilidades de contribuição/vantagens para a prática docente do profissional de línguas com o uso da Linguística de Córpus: aprendizagem da



09 A 13 DE NOVEMBRO DE 2020

frequência das palavras; colocações; frases e fragmentos lexicais; gramática; exemplos autênticos; observação, reflexão e conhecimento do léxico em uso; abordagens descritivas; palavras novas; corpora especializado/específico e revolução digital.

Metodologia

Gil (2017) define pesquisa como um procedimento sistemático e racional que tem como objetivo principal a busca de respostas a determinados problemas. Requer-se uma pesquisa quando as informações dispostas ainda não são suficientes para a resposta a um determinado problema ou quando essas informações estão desordenadas não podendo ser adequadamente aplicadas.

Para que uma pesquisa seja desenvolvida, servimo-nos dos conhecimentos disponíveis e da utilização criteriosa de métodos, técnicas, bem como de outros procedimentos científicos. Na pesquisa, é desenvolvido um processo que vai desde a formulação dos problemas à apresentação satisfatória dos resultados. Existem muitas razões para a realização de uma pesquisa; intelectualmente, pelo próprio desejo de conhecer e no aspecto prático pelo desejo de fazer algo mais eficiente e eficaz.

Essa pesquisa é de abordagem quantitativa cujo objetivo maior é a descrição. Por pesquisa quantitativa, ressaltamos que está enraizada no pensamento lógico positivista cuja ênfase se dá nos atributos mensuráveis da experiência humana. Parte da localização quantitativa de determinados conceitos e busca compreender a totalidade do fenômeno a partir de ideias acerca da forma pela qual os conceitos estão relacionados. Utiliza procedimentos estruturados e instrumentos formais para a coleta de dados mediante condições de controle que capturem o contexto na totalidade. Em nenhuma hipótese, abre mão da objetividade na coleta e análise dos dados que por sua vez se dão por meio de procedimentos estatísticos.

Com relação ao seu caráter descritivo, tem como objetivo primordial descrever características de um determinado fenômeno estabelecendo relações entre variáveis a partir de técnicas padronizadas de coleta de dados. A pesquisa descritiva exige do investigador uma série de informações sobre o que se deseja pesquisar para descrever fatos e fenômenos de uma determinada realidade. Nela, são obtidos dados que podem ser analisados estatisticamente. As variáveis podem ser quantificadas e é possível realizar correlações entre procedimentos estatísticos tornando-se possível conhecer a margem de erro dos resultados obtidos na amostragem.

Material

O material empreendido na pesquisa se desdobra em textos escritos por alunos de Ensino Médio para um processo seletivo. Os textos que compõem o *cópus* de estudo totalizam 51 redações. Para o presente estudo, foi realizado um recorte de dez redações.

Para a análise do *cópus*, utilizamos o *software WordSmith Tools*, eficaz nos processos de descrição e operacionalização. Na investigação da linguagem, o emprego de computadores proporciona resultados consistentes, contando palavras e identificando as ocorrências de forma confiável.



09 A 13 DE NOVEMBRO DE 2020

O *WordSmith Tools* foi idealizado por Mike Scott e publicado pela Universidade de Oxford. Põe à disposição do analista inúmeros recursos para a análise de vários aspectos da linguagem, dentre eles a composição do léxico, o tema dos textos e a organização composicional e retórica de gêneros discursivos. O *software* é composto pelas seguintes ferramentas: *Wordlist*, *Key Words*, *Cconcord* e possui os seguintes utilitários: *Renamer*, *text converter* *splitter e viewer*. Cada ferramenta com ajustes e funções determinados. Trabalha com os princípios abstratos de ocorrência, recorrência e coocorrência.

Para que fosse utilizada em consonância com as redações e com o *WordSmith Tools*, recorreremos à uma lista de palavras, obtida do cópulus do Português popular brasileiro escrito, denominada CorPop (PASQUALINI, 2018). Seus textos foram selecionados com base no nível de letramento médio dos leitores do país. Uma vez que a lista CorPop não encontra-se lematizada, foi necessário ampliá-la em forma de lista deslematizada. Como produto, obtivemos uma nova lista com 5.237 palavras, acrescentando as suas variações e flexões. Essa lista nos serviu de base para identificarmos as palavras mais formais nas redações que compõem o cópulus. Finalmente, utilizamos o site <https://www.socscistatistics.com/tests/spearman/default2.aspx> para estabelecermos os cálculos de correlação entre as notas dos corretores e o perfil lexical.

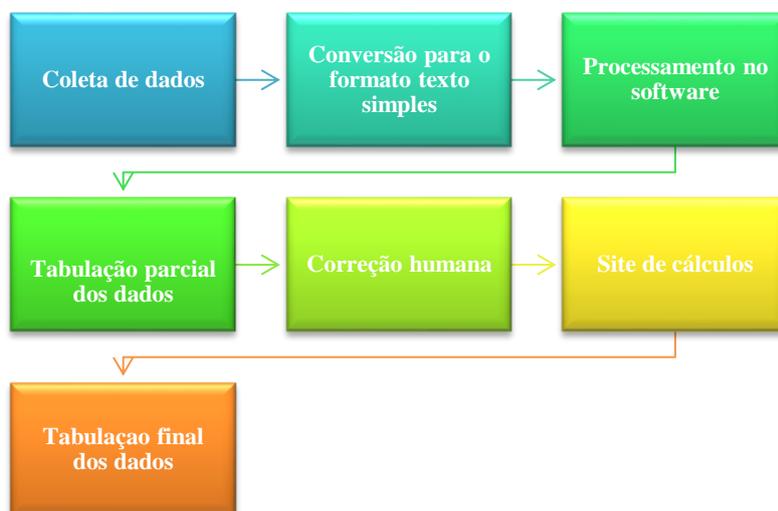
Método

A pesquisa foi dividida nas seguintes etapas: a coleta dos textos, a conversão do texto para o formato txt (texto simples), o processamento dos textos no software, a tabulação parcial dos dados, a correção humana, a correlação entre o perfil lexical e a nota dos corretores e a tabulação final dos dados.



09 A 13 DE NOVEMBRO DE 2020

Figura 1 - Fluxograma com as etapas do método dessa pesquisa



Fonte: Dados da presente pesquisa.

O primeiro passo foi a coleta das redações realizadas por alunos do Ensino Médio. Os textos, ao serem coletados, foram digitalizados e digitados. Em seguida, convertidos para o formato txt (texto simples), sem formatação, exigência do software para que o corpúss fosse processado.

A partir da lista de palavras deslematizadas, tivemos um repertório lexical abrangente que ao ser colocado como lista-base nos permitiu identificar palavras mais formais – aquelas que não constavam na lista-base. O intuito era identificar se o perfil lexical que tende à formalidade tem impacto na nota atribuída pelos corretores.

As redações foram corrigidas por 10 professores. Os professores tiveram a liberdade de corrigir as redações a partir de seus próprios critérios. Antes dessa correção, foi repassado a eles um formulário de avaliação contendo algumas perguntas específicas que os habilitavam ou não para a participação na pesquisa. Este formulário foi criado no *Google Forms*, dividido em duas seções. A primeira seção continha perguntas relacionadas à formação acadêmica e à área de atuação dos professores convidados. Já a segunda seção, continha as dez redações a serem corrigidas.

Após traçarmos o perfil lexical das redações e os professores finalizarem as correções e atribuição de notas, procedemos ao cálculo de correlação, estabelecendo uma comparação entre as notas dos corretores e a quantidade de palavras de cada redação que não constava na lista ampliada.



09 A 13 DE NOVEMBRO DE 2020

Resultados e Análise dos Dados

Pensando acerca da face quantitativa da linguagem humana e na infinidade de frequências dos fenômenos discursivos, torna-se justa a adoção de uma ótica estatística. Nesse enfoque, a língua pode ser vista como uma fonte de extração de dados. Biderman (2001) ressalta que o fenômeno linguístico constitui um dos mais significativos para um tratamento quantitativo.

A correção das redações realizadas pelos professores demonstrou notas variadas, como já era esperado. Sendo assim, estabelecemos a mediana das notas para o recorte de 10 (dez) redações corrigidas. As medianas estabelecidas nos serviram de índice para a comparação. Apresentamos, a seguir, as medianas das notas atribuídas pelos professores a cada redação.

Tabela 2 – Notas atribuídas às redações na correção humana

	REDAÇÃO									
	1	2	3	4	5	6	7	8	9	10
Corretor 1	10	8	1	3	6	7	8	7	9	8
Corretor 2	2	4	0	7	9	0	0	1	10	7
Corretor 3	7	6	4	3	6	3	6	0	8	6
Corretor 4	5	5	4	4	8	2	5	3	9	8
Corretor 5	6	8	6	5	7	5	6	4	7	9
Corretor 6	7	5	3	6	10	1	6	4	10	7
Corretor 7	6	4	4	7	8	0	6	2	8	4
Corretor 8	6	4	1	7	9	2	6	2	10	4
Corretor 9	6	3	2	7	8	4	7	5	9	9
Corretor 10	3	3	1	3	5	1	4	1	8	4
mediana	6	4,5	2,5	5,5	8	2	6	2,5	9	7

Fonte: Dados da presente pesquisa.

Sabemos que que uma redação não se constitui somente de vocabulário, critérios de coesão, coerência, pertinência ao tema, dentre outras competências devem ser analisados. No entanto, para a pesquisa em pauta nos detivemos na análise da variabilidade de vocabulário dos textos.

Como pode ser visto, o menor valor de mediana foi para a redação de número 6. E o maior valor foi para a redação de número 9. As redações foram digitadas fielmente à escrita original.

Apresentamos a redação com menor mediana (2) estabelecida pelas correções docentes.

Sim, mas no Brasil do Estados tenho r\$ 1 de deficiência é surdos todos primeiro língua portuguesa L1 faltar depois vai faz L2 mais pessoais e surdos se faltar coisa, na escola tem alguns comunicação com profissões mas se exemplo está não ter de interpretar faltar si difícil o que conseguir libras profissões só escrever mostrar do surdo estar não entender nada



09 A 13 DE NOVEMBRO DE 2020

mais ruim. Mais melhor só primeira escola bilíngue, em bilíngue é bom no libras como primeira, porque dentro tem coisa palavrão pronto todos também educação bilíngue mas bilíngue escola o quê de surdos ou surdas ter profissões é educação todos. No Brasil todos sentindo concordo, história criar primeiro onde em Rio de Janeiro, mas libras de primeiro criar onde é país no França.

Analisando a redação acima, percebemos desvios ortográficos, de concordância e a repetição de palavras. Em um nível lexical, ocorreu a ascensão de vocábulos que designam um campo lexical específico, nesse caso: bilíngue, língua portuguesa, escola, interpretar, Libras, escrever, palavrão, educação, etc. É importante ressaltar que chamamos de campo lexical o conjunto de palavras de uma língua que estão relacionadas por afinidades conceituais. As escolhas lexicais na redação acima foram feitas para atender a um determinado propósito comunicativo. Nesse caso, os termos escolhidos, embora não de forma ampla, induzem-nos a pensar acerca do tópico pré-determinado. Ainda que o conhecimento lexical de um aluno seja vasto, é necessário que o pensamento seja organizado de tal modo que ocorra um emprego coerente das estruturas linguísticas na modalidade escrita do texto.

Sabendo que a coesão lexical ocorre quando a intenção é evitar repetições; substituindo termos por pronominalizações, sinônimos ou expressões equivalentes, fica claro na análise desvios nesse aspecto. Determinados termos foram por vezes repetidos, como por exemplo, a palavra bilíngue, profissões, escola, Libras.

O aumento de insumo lexical que o aprendiz tem pode impactar em seu desempenho em uma redação, por exemplo – a depender da exposição e emprego consciente do léxico. É oportuno que esses indivíduos aumentem seu repertório lexical ao longo da formação no que se refere às habilidades de compreensão e produção porquanto sua produção linguística tende a ser avaliada, especialmente no contexto escrito.

Apresentamos a redação com maior mediana (9) estabelecida pelas correções docentes.

Ao analisar a questão da formação educacional de surdos no Brasil fica evidente que a sociedade não herdou de Roma somente o modelo do Código Civil, mas também a ideia de marginalização dos portadores de deficiências, visto que estes ainda encontram dificuldades para inclusão social.

Mesmo que, hoje, lhe seja assegurado o direito à vida, outras questões ainda são problemáticas. Tal defasagem ocorre por conta da falta de adaptação da sociedade, desde o âmbito escolar ao profissional. Sobre o primeiro, apesar das leis que garantem o direito à educação, muitos lugares apresentam despreparo para fornecê-la, isto é, ainda é comum que professores de escolas convencionais não saibam comunicar-se através da libras. Assim como nas empresas onde dificilmente existem funcionários capazes de realizar uma entrevista de emprego com portadores da deficiência. Além disso, por diversas vezes, o acesso aos lugares que oferecem tais diferenciais se torna problemático para a independência do indivíduo, pois é quase inexistente o número de motoristas de transportes públicos e privados, como táxis que sabem como orientar ou até mesmo



09 A 13 DE NOVEMBRO DE 2020

entendê-lo em caso de dúvida. Então, tem se todos esses exemplos como afirmação da ausência da inclusão e não somente como justificativa tem se a herança histórica, mas também a direção que as relações sociais tem tomado, como exemplifica Bauman em "Modernidade Líquida" onde constrói-se uma sociedade cada vez mais individual. No entanto, apenas justificar e não corrigir o problema é um equívoco. Deve-se em síntese, alterar o meio social em que a população está inserida, logo, o ministério da Educação deve promover o ensino obrigatório de Libras através das escolas para que a aprendizagem da língua não seja mais vista como obrigação exclusiva dos familiares do indivíduo e também cursos obrigatórios nos ambientes profissionais para que a comunicação alcance as mesmas proporções de fala que o português.

Analisando a redação acima, em um nível lexical, percebe-se a variabilidade de palavras utilizadas. Poucas foram as repetições, o que demonstra um alto grau de insumo lexical por parte do escritor. As escolhas lexicais também atenderam à finalidade comunicativa estabelecida: os vocábulos se deram dentro da proposta do texto, com alto índice de diversificação, não demonstrando inadequações vocabulares. Percebe-se considerável precisão vocabular, apropriada ao texto, o que interfere sobremaneira na inteligibilidade da produção. As redações foram analisadas de modo individual com o intuito de identificarmos a variabilidade lexical utilizada nos textos.

Tabela 1 - Escala de correlação

<i>Valor de p</i>	<i>Interpretação</i>
0.0 a 0.19	Uma correlação bem fraca
0.20 a 0.39	Uma correlação fraca
0.40 a 0.69	Uma correlação moderada
0.70 a 0.89	Uma correlação forte
0.90 a 1.00	Uma correlação muito forte

Fonte: Larson e Farber (2015).

Seguindo a escala acima, entendemos que se obtivermos o valor 1.0, temos uma correlação perfeita, isto indica que quanto mais palavras fora da lista ampliada, maior a nota. A correlação varia de muito fraca até muito forte.

Vejamos a correlação entre as notas do corretor e os types da redação:

Quadro 2 – Correlação entre as notas do corretor e os types da redação

Corretor 1	Corretor 2	Corretor 3	Corretor 4	Corretor 5	Corretor 6	Corretor 7	Corretor 8	Corretor 9	Corretor 10
0.23	0.44	0.60	0.72	0.67	0.52	0.52	0.55	0.49	0.79

Fonte: Dados da presente pesquisa.

A partir da escala exposta, é possível observar que para o Corretor 1, as palavras mais formais tiveram uma influência fraca. No entanto, para os Corretores 2, 3, 5, 6, 7, 8 e 9, a influência foi moderada, o que reforça a ideia de que quanto maior a quantidade de palavras



09 A 13 DE NOVEMBRO DE 2020

mais formais, maior é a nota atribuída. Para os Corretores 4, e 10, a correlação é forte, indicando que a nota sofreu influência positiva conforme a redação apresentava palavras formais.

Baseando-nos no pequeno *cópus* de redações e no pequeno grupo de participantes, entendemos que existe uma correlação positiva entre a presença de vocabulário formal e a nota atribuída pelos corretores para oito corretores. Portanto, para os professores participantes da pesquisa, a qualidade do vocabulário tem um impacto importante na redação, variando de moderado à forte.

Tal assertiva pode ser considerada pelos professores de Língua Portuguesa e de redação na preparação e execução das aulas.

Considerações finais

A partir das análises realizadas, partindo, sobretudo, do princípio de que o professor pode direcionar seus alunos na aprendizagem do vocabulário, é preciso atentar-se para as palavras que se têm trabalhado em sala.

Conforme demonstrado, na investigação da linguagem, o emprego de um *software* de processamento linguístico e uma lista de palavras frequentes deslematizada proporciona resultados consistentes, identificando as ocorrências de forma confiável, possibilitando que novos fatos sejam descobertos e opiniões e crenças sejam contestadas.

Identificamos que existe um efeito positivo na nota atribuída pelos professores quando estão presentes palavras mais formais nas redações. Em vista das ideias apresentadas, torna-se relevante para professores e alunos refletirem acerca do conjunto de palavras utilizadas. Esse vocabulário faz parte do léxico individual que também forma um léxico global que compreende o conjunto de palavras recorrentes num momento sociohistórico específico. Dificuldades de se produzir textos podem ser decorrentes da falta de conhecimento de um vocabulário amplo.

Referências

BERBER SARDINHA, T. Como usar a linguística de corpus no ensino de língua estrangeira: por uma linguística de corpus educacional brasileira. *In*: VIANA, V.; TAGNIN, S.E.O. (org.) **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010

BERBER SARDINHA, T. Linguística de Corpus. *In*: GONÇALVES, A. V.; GÓIS, M. L. S. (org.). **Ciências da linguagem: o fazer científico?** v. 1, Campinas: Mercado de Letras, 2012. p. 321-347.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics: investigating language structure and use**. Cambridge: Cambridge University Press, 1998.

BIBER, D.; REPPEN, R. Introduction. *In*: BIBER, D.; REPPEN, R. (ed.). **The Cambridge handbook of English Corpus Linguistics**. Cambridge: Cambridge University Press. 2015. p. 1-8.



09 A 13 DE NOVEMBRO DE 2020

BIDERMAN, M. T. C. **Léxico e vocabulário fundamental**. *Alfa*, São Paulo, p. 27-46, 1996.

BIDERMAN, M. T. C. **Teoria linguística: leitura e crítica**. 2. ed. São Paulo: Martins Fontes, 2001.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2017.

LARSON, R.; FARBER B. **Estatística aplicada**. 6. ed. São Paulo: Pearson, 2015.

LEECH, G. N. **Corpus and theories of linguistic performance**. Berlin: Mouton de Gruyter, 1992.

LEECH, G. N. **The state of art in Corpus Linguistics**. London: Longman, 1991.

MOORE, D. Uma didática da alternância para aprender melhor? In: PRADO, C.; CUNHA, J. C. (orgs.) **Língua materna e língua estrangeira na escola**. O exemplo da Bivalência. Belo Horizonte: Autêntica, 2003.

NOVODVORSKI, A.; FINATTO, M. J. B. **Linguística de Córpus no Brasil: uma aventura mais do que adequada**. *Letras & Letras*, Uberlândia, v. 30, n. 2, p. 7-16, dez. 2014.

PASQUALINI, B. F. **Corpop: um corpus de referência do português popular escrito do Brasil**. Universidade Federal do Rio Grande do Sul. Porto Alegre. 2018.

SARMENTO, S. **Linguística de Córpus: histórico, metodologia, campos de aplicação**. *Trama*, Marechal Cândido Rondon, v. 6, n. 2, p. 87-107, 2010.

SINCLAIR, J. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.