

## Vocabulário escrito de estudantes e de materiais didáticos do Ensino Fundamental nas redes públicas municipal e estadual de educação<sup>1</sup>

### Written vocabulary of Middle school students and textbooks in State and Municipal public schools

*Bruna Rodrigues da SILVA\**

---

**RESUMO:** Este artigo relata o primeiro estudo piloto de pesquisa de mestrado na área dos estudos do léxico, e se dedica a identificar o perfil de vocabulário escrito de estudantes do Ensino Fundamental. O objetivo é descrever, de maneira inicial, o vocabulário de alunos do Ensino Fundamental nas redes públicas municipal e estadual de educação. O vocabulário dos estudantes é examinado conforme manifestado em suas produções escritas e é contrastado com o de textos de materiais didáticos por eles utilizados. Com o apoio da estatística linguística (BIDERMAN, 1978, 1998) e da Linguística de Corpus (BERBER SARDINHA, 2004), a ideia é gerar contrastes e diferenciações entre esses dois universos. Foram examinados oito textos de alunos, dois textos de livros didáticos, além de um texto de apoio, utilizado junto com a proposta de produção textual. Os elementos observados, nas três fontes textuais, foram o número de tokens e de types, a frequência das palavras,

---

**ABSTRACT:** This paper is an account of the first pilot study that was carried out in support of the Master's-level research in Lexical Studies. The investigation looks at the written vocabulary of Middle school students in municipal and state public schools. Students' written production will be assessed and will be contrasted with the vocabulary of the textbooks they use in class. Relying on the contributions of Linguistic Statistics (BIDERMAN, 1978, 1998) and Corpus Linguistics (BERBER SARDINHA, 2004), we compare and contrast these two universes. Eight writing samples written by students and two textbook texts, including an issue topic were used. The following items were studied in the three textual sources: number of tokens and types, the frequency of the words, the difference between the number of occurrences of lexical and grammatical words, the vocabulary richness and the hapax legomena. Results show

---

<sup>1</sup>Este artigo engloba e expande o estudo piloto I, realizado para a dissertação de mestrado intitulada: **Vocabulário escrito de estudantes de Ensino Fundamental de Escolas Públicas**. Dados apresentados aqui integram as discussões do artigo: TCACENCO, SILVA, FINATTO. **Acessibilidade textual e terminológica:** conquistas recentes, pesquisas em andamento e novas perspectivas (2019), disponível [aqui](#).

\* Mestranda do Programa de Pós-Graduação em Letras da UFRGS, professora da rede pública de educação. ORCID: <https://orcid.org/0000-0002-0597-3624>. [thu\\_du@hotmail.com](mailto:thu_du@hotmail.com)

---

a diferença entre as ocorrências de palavras lexicais e gramaticais, o hapax legomena e a riqueza vocabular. Os resultados indicam que a riqueza vocabular dos textos dos alunos não varia entre as duas escolas e que não há diversidade de palavras. Ao analisar as vinte palavras mais frequentes, nota-se que as que mais aparecem também estão presentes no texto de apoio, contudo, o repertório vocabular dos textos dos livros didáticos não é utilizado pelos estudantes. O estudo evidencia a necessidade de maior ponderação quanto ao número de textos a serem observados nos diferentes segmentos do corpus sob exame.

**PALAVRAS-CHAVE:** Vocabulário escrito. Redações escolares. Língua Portuguesa. Ensino Fundamental. Escola pública.

that students' vocabulary richness is similar in the two schools and that there is no diversity of words. When analyzing the twenty most frequent words, we noted that the most frequent words are also present in the issue topic. However, the vocabulary presented in the textbooks is not used by students. The study shows the importance to assess the number of texts to be studied in the different parts of the corpus under study.

**KEYWORDS:** Written vocabulary. School essays. Portuguese language. Middle school. Public school.

---

## 1 Introdução

Este trabalho relata resultados do primeiro estudo piloto relacionado a uma pesquisa de mestrado que se insere no âmbito dos estudos lexicológicos, com apoio da estatística linguística (BIDERMAN, 1978, 1998) e da Linguística de Corpus como abordagem (BERBER SARDINHA, 2004). O foco da dissertação como um todo é a descrição e a análise contrastiva do perfil do vocabulário escrito empregado por estudantes de Ensino Fundamental das redes públicas municipal (Porto Alegre) e estadual (Rio Grande do Sul) de educação.

A produção escrita dos estudantes, objeto de estudo principal, é contrastada com o repertório vocabular dos materiais didáticos por eles utilizados. A partir de uma amostra de redações produzidas pelos estudantes, que integram um acervo pessoal da professora de Português desses alunos, foi organizado um *corpus* de estudo, composto por: a) 54 redações (sendo 94 dos alunos do 8º ano/C20 e 60 dos alunos do 6ºano/B30)

de duas escolas diferentes, uma estadual e outra municipal; b) 5 textos de apoio - aos quais os alunos foram expostos antes de escreverem os seus – e registros de motivação para produção escrita; e, c) 2 textos dos livros didáticos de Português utilizados pelos estudantes.

O objetivo principal da pesquisa como um todo é detalhar os aspectos léxico-estatísticos do *corpus* de estudo, para, a partir deles, descrever o perfil do vocabulário escrito dos estudantes de Ensino Fundamental das redes públicas. Para tanto, serão necessárias comparações entre os dois grupos e com os elementos de contraste: textos de apoio às propostas de produções textuais, as próprias propostas e textos dos livros didáticos.

Para o estudo piloto, aqui relatado, em meio ao acervo de dados de pesquisa, foram selecionados como amostra: quatro produções de cada grupo de estudantes, o respectivo texto gerador, a proposta de produção e um texto de um livro didático de cada escola, o qual tenha sido trabalhado com os estudantes antes de realizarem suas produções escritas. Os elementos observados, neste primeiro momento, foram o número total de palavras (*tokens*), a variedade de palavras empregadas (*types*), os graus de frequência das palavras, a proporção de emprego de palavras lexicais e gramaticais, a riqueza vocabular, (*types/tokens ratio*) e, por fim, as palavras de frequência única (*hapax legomena*). Os resultados indicaram que a medida de riqueza vocabular dos textos dos alunos não tende a variar entre os dois grupos de estudantes e que a diversidade de palavras parece estável entre os jovens das duas escolas. Além disso, ao contrastar o vocabulário dos estudantes e o do texto de apoio, nota-se que as palavras mais empregadas estão presentes ou no texto de apoio ou na proposta de produção textual. Quanto aos textos didáticos, houve coincidência apenas entre o repertório de palavras gramaticais. Os indicativos do trabalho apontam rumos possíveis da pesquisa em maior escala e são ponderados os números de textos necessários em cada segmento do *corpus* de estudo.

O texto que segue, que detalha os resultados acima mencionados, está organizado da seguinte maneira: a próxima seção relata experiências anteriores associadas a pesquisa como um todo, na qual se insere o estudo piloto. A seção seguinte aborda nosso referencial teórico e prático, enquanto a seção quatro traz nossas hipóteses iniciais. A partir dessas hipóteses, relatamos o experimento realizado em sala de aula, que deu origem ao pequeno *corpus* desse primeiro estudo, na seção cinco. Além disso, ainda na seção cinco, descrevemos nosso primeiro estudo piloto, os contrastes produzidos, os resultados obtidos e as discussões realizadas. Na sequência retornamos às nossas hipóteses iniciais, na seção seis, e, por fim, na seção sete, a partir desse primeiro estudo piloto, apresentamos diferentes aspectos que precisam ser considerados na pesquisa em maior escala, bem como na dissertação como um todo.

## **2 Antecedentes, foco da pesquisa e comunidades envolvidas**

Na época da Graduação em Letras, concluída no ano de 2011, por meio da experiência com a Iniciação Científica, estudamos sobre temáticas relacionadas aos estudos de léxico e de vocabulário. Com tal experiência, iniciamos a prática docente, como professora de Língua Portuguesa, sempre em escolas públicas, em diferentes municípios do Estado do Rio Grande do Sul. Hoje, atuamos em uma escola do Município de Porto Alegre e em outra do Estado do Rio Grande do Sul, ambas situadas na cidade de Porto Alegre, em um bairro considerado periférico.

Nesses dois ambientes, pertencentes a administrações diferentes, vivenciamos as várias questões que englobam a escola pública. A partir disso, percebemos como a escola é importante e pode fazer a diferença na vida dos alunos. Sobretudo para quem, muitas vezes, não tem acesso e nem oportunidades fora desse ambiente escolar.

Nesse contexto, algo que sempre nos intrigou é a dificuldade de leitura, de compreensão e de interpretação dos alunos dessas escolas. Além disso, subjetivamente, sempre reparamos na diferença de desempenho e de aproveitamento entre alunos da

escola estadual e alunos da escola municipal, considerando que cada escola é regida por princípios e ordenamentos jurídicos e pedagógicos diferentes.

A constatação das dificuldades ultrapassa a disciplina de Língua Portuguesa, sendo também percebida e questionada por muitos colegas, professores de outras áreas, que atuam nas mesmas redes escolares. Segundo Nunes (2014) “[...] nosso dever, na formação dos alunos de ensino fundamental e médio, é contribuir para tornar aqueles que já falam a língua portuguesa em contextos menos controlados falantes mais proficientes” (NUNES, 2014, p.53).

Em função disso, em meio a um amplo espectro de dificuldades, de diferentes naturezas, tais como falhas na ortografia, na pontuação e outros problemas de domínio incompleto da norma escrita culta, cremos que a base para entender as dificuldades específicas dessa população estudantil pode estar no equacionamento do tipo de ensino de língua materna oferecido. É preciso também um exame dos materiais didáticos utilizados, bem como do modo de apresentação e desenvolvimento das atividades pedagógicas. Importa, por fim, compreender como esses elementos dialogam, por exemplo, entre si e com as condições e perfis dos alunos.

Em meio à produção escrita escolar, voltamos nosso olhar para os materiais didáticos de Português disponíveis nas escolas em que lecionamos atualmente: coleção Tecendo Linguagens, de Oliveira...[et al] (2015), e coleção Português: Linguagens, de Cereja e Magalhães (2014), referentes ao 8º ano.

Ao longo da nossa experiência como professora, em um intervalo de tempo de cerca de seis anos, iniciamos a construção de um acervo pessoal de textos de alunos. Esse acervo, hoje, fornece o principal material para o nosso *corpus* de pesquisa. Para tanto, cada texto foi devidamente desidentificado e associado a registros que são: a) texto de apoio para produção da redação; b) avaliação, conceito e/ou comentário de cada texto por parte da professora; e c) registro dos textos dos livros didáticos trabalhados antes da produção escrita.

Feita essa apresentação de nossas experiências anteriores e de como nos reaproximamos da pesquisa acadêmica, passamos à apresentação de alguns estudos prévios relacionados ao nosso.

## 2.1 Trabalhos relacionados

Conforme pesquisamos no acervo da UFRGS, sistema Sabi<sup>2</sup>, em 2019, a pesquisa com produções escritas de alunos de Ensino Fundamental ainda é relativamente pouco explorada. Em pesquisa fora desse acervo, avançando para outros estados do Brasil, ao que parece, a grande maioria dos estudos realizados com textos ocupa-se de redações de vestibular ou de Enem ou, ainda, de textos de estudantes de Ensino Médio (que, por consequência, irão prestar vestibular e/ou Enem).

No âmbito do nosso Estado e Universidade, Endruweit (2000), por exemplo, pesquisou sobre a evolução das temáticas da redação do vestibular da UFRGS. A autora apresentou, entre diferentes dados, a pressão que esse assunto sempre exerceu sobre os alunos do Ensino Médio, bem como trouxe uma análise das características de uma redação que possa ser considerada nota 10 nesse concurso.

Por outro lado, considerando-se um cenário nacional, Luiz (2018) investigou a qualidade e a relevância social das redações estilo ENEM produzidas por alunos de ensino médio de uma escola particular de Uberlândia (MG). Além desses, é possível encontrar muitos outros trabalhos centrados nos mesmos tipos de produções textuais: redação de vestibular e/ou ENEM.

Conforme entendemos, pouca atenção se dá aos textos de alunos de Ensino Fundamental, que parecem ainda pouco explorados de maneira científica. Portanto, nossa ideia é, de alguma maneira, contribuir para diminuição dessa lacuna.

---

<sup>2</sup> Catálogo *on-line* da Universidade Federal do Rio Grande do Sul, que inclui artigos, trabalhos de pós-graduação, trabalhos de conclusão de curso, entre outros tipos de produção intelectual, de toda comunidade universitária. Disponível em: <https://sabi.ufrgs.br>

A seguir são apresentados os dois cenários educacionais que servem de pano de fundo para nossa pesquisa: as escolas públicas do Estado do Rio Grande do Sul, nos níveis estadual e municipal.

## 2.2 Comunidades envolvidas

As escolas cujas vivências pretendemos aproveitar para o todo da nossa pesquisa situam-se na Zona Norte de Porto Alegre, capital do Rio Grande do Sul, ambas no bairro Sarandi, cerca de 3km de distância uma da outra. O fato de uma escola ser estadual e, a outra, municipal, é algo a ser inserido no cenário de pesquisa, considerando-se duas realidades de ensino público, postas em contraste.

A primeira escola que fornece dados de observação contextual é a Escola Estadual de Ensino Fundamental Dr. Ferreira de Abreu, Escola de Ensino Fundamental administrada pelo Governo do Estado do Rio Grande do Sul, fundada em 1943. Lecionamos nessa escola, nas turmas de 6º a 9º ano do Ensino Fundamental, desde o ano de 2013. Atualmente a escola abriga 430 alunos.

A segunda é a Escola Municipal de Ensino Fundamental Décio Martins Costa, escola de Ensino Fundamental administrada pela Prefeitura de Porto Alegre, fundada em 1963. Lecionamos nessa escola, nas turmas de primeiro ano e de segundo ano do terceiro ciclo, respectivamente, C10 e C20, desde o ano de 2016. Atualmente tem 924 alunos.

A seção seguinte apresenta, em linhas gerais, os referenciais teóricos e metodológicos da nossa pesquisa como um todo.

## 3 Referencial teórico e metodológico

Conforme mencionado na introdução deste artigo, a descrição e as análises com nosso *corpus* são feitos à luz de referenciais teórico-metodológicos da Linguística de

*Corpus* (LC), introduzida no nosso país pelos trabalhos de Beber Sardinha (2004). Essa é uma subárea da Linguística Aplicada, que se dedica ao estudo da língua em grandes conjuntos de textos autênticos, denominados *corpus/corpora*, contando com o apoio de ferramentas informatizadas. Há, na LC, a proposta do tratamento estatístico para o que oferecem os *corpora*. Um *corpus*, na acepção da LC, é uma reunião de textos em formato digital, criteriosamente reunidos, que serve para representar um dado estado de uso de uma língua (BERBER SARDINHA, 2004, p. 16-17).

Na nossa pesquisa, lidamos com três acervos: 1) os textos dos alunos, 2) os textos de apoio e/ou registros de motivação às propostas de produção e 3) os textos dos livros didáticos de Português. Nosso *corpus* de estudo é composto pelo acervo 1. Os demais acervos correspondem a elementos de contraste, que nos permitem situar o perfil de vocabulário empregado pelos estudantes. Todos esses materiais foram digitalizados para que fosse possível a condução da pesquisa e a utilização de ferramentas computacionais.

Além da LC, outro referencial importante para nossa pesquisa são os trabalhos pioneiros de M.T.C. Biderman (1978, 1998 e outros). Essa é uma fonte importante, especialmente em relação à estatística linguística. Essa autora foi uma das primeiras a evidenciar a face quantitativa da linguagem. Segundo ela, a estatística linguística constata a estabilidade dos e torna possível a previsibilidade dos símbolos linguísticos (BIDERMANN, 1998).

Como tratamos de léxico, cuja unidade tradicional é denominada “palavra”, é preciso escolher também um ponto de referência para nossa pesquisa. Sobre o que exatamente seja uma palavra, existem várias concepções e, conforme Finatto (2009) cada ponto de vista parece indicar um objeto diferente. Neste estudo, cujo encaminhamento é restrito à palavra escrita, acreditamos que a melhor opção para definir o termo seja a de seguir Bisognin (2009), para quem palavra é “(...) toda unidade lingüística



mínima que pode constituir significado, delimitada na escrita por dois espaços em branco e/ou sinal de pontuação” (BISOGNIN, 2009, p. 25).

Inspiradas em Basilio (2004), consideramos o vocabulário como uma espécie de banco de dados da língua, um depósito de elementos de designação, que fornece unidades básicas e que é capaz de se expandir, originando novas unidades. Em que pese essa escolha sobre o escopo do termo vocabulário, há diferentes conceituações possíveis. Conforme Bisognin (2009): “No sentido mais comum, léxico é o repertório de palavras de uma língua.” (BISOGNIN, 2009, p. 26). Há diferentes pontos de vista e diferentes ponderações conceituais e terminológicas, que, pela natureza deste trabalho, não serão desenvolvidos aqui.

Analisaremos elementos lexicais (que têm um sentido associado a fatos, fenômenos, processos, propriedades, modalidades, com conteúdo nocional mais específico) e elementos gramaticais (com sentido dependente da combinatória em que se encontram/do funcionamento da língua). Há concepções que consideram o léxico apenas como o conjunto das palavras que possuem um sentido autônomo na língua, como os verbos e substantivos, desconsiderando elementos estritamente gramaticais, como os artigos e as preposições (SILVA; FINATTO, 2009). Entretanto, analisaremos aqui os diferentes itens do vocabulário da língua (tanto ‘mulher’ ou ‘viajar’, quanto ‘de’ ou ‘mas’) que, embora tenham características distintas, fazem parte do léxico aqui considerado.

Biderman (1998) declara que “Com base nos resultados da Estatística Léxica ou Léxico-estatística podemos afirmar também que a frequência é uma característica típica da palavra” (BIDERMAN, 1998, p. 162). Em função disso, interessou-nos observar os diferentes padrões de frequência das palavras que aparecem no *corpus* de estudo para, a partir disso, caracterizá-las de acordo com a sua presença e função nos textos, contrastando os diferentes resultados entre si.

A seguir são apresentadas as hipóteses iniciais do nosso primeiro estudo piloto. Os resultados nele obtidos deverão subsidiar o estudo piloto número dois.

#### **4 Hipóteses iniciais**

A partir de uma amostra inicial de textos, nossa primeira hipótese, baseada em nossa experiência docente, é de que os alunos tenderiam a exibir um vocabulário repetitivo ao escreverem seus textos, isto é, a tendência seria de que a riqueza lexical seja limitada. Imagina-se que isso ocorra porque os jovens normalmente não releem suas produções antes de entregá-las e nem costumam reescrevê-las. Aqui há o pressuposto de que se o estudante reler o que escreve, tenderá a diversificar o vocabulário empregado.

Outra aposta é de que o léxico mais frequentemente utilizado pelos jovens é, provavelmente, aquele que também foi utilizado pelos professores em sala de aula, ao explicar a proposta de trabalho, além daquele presente em textos de apoio ou textos lidos e trabalhados recentemente, ligados à temática da redação proposta.

Além disso, embora os dois públicos envolvidos sejam de escolas públicas, localizadas na mesma região da capital, logo que iniciamos o trabalho na Rede Municipal, percebemos que havia grande diferença entre estes dois cenários de ensino público e, principalmente, entre os sujeitos dessa aprendizagem: os alunos. Subjetivamente, enquanto professoras das duas realidades, em paralelo, e também de acordo com colegas docentes, acreditamos que os alunos da escola municipal (apesar desta rede dispor de professores especializados e com maior formação, laboratórios de aprendizagem, diversos projetos, entre outros) tendem a apresentar maior dificuldade em relação à leitura, à compreensão e à interpretação. Exibem também falhas de ortografia, de formação de palavras e de frases, de concordância e de conjugação de verbos, entre outros.

Dessa forma, a última hipótese é de que os alunos das duas redes educacionais apresentam diferentes desempenhos e de que os da Rede Municipal apresentam dificuldades específicas, diferentes das dificuldades do grupo da Rede Estadual. Neste particular, acreditamos que essa hipótese pode ser verificada no âmbito do vocabulário exibido via produção escrita dos estudantes. Assim, o exame do vocabulário das produções escritas dos alunos seria uma especificidade capaz de revelar pistas que nos auxiliem a explicar e a compreender possíveis diferenças entre os desempenhos dos dois públicos.

## 5 Experimento realizado

Os textos dos alunos são peça principal sob exame, tanto neste primeiro estudo, quanto na dissertação como um todo. Em função disso, nossa opção para iniciar a exploração do *corpus* de estudo foi a de selecionar cerca de 20% do total de textos de uma dada unidade de ensino trabalhada com todas as turmas de 8º ano/C20 das duas escolas. Também foram observados o respectivo texto de apoio para a produção da redação e um texto de cada um dos livros didáticos de Português utilizados nestas duas escolas.

Nesta seção são descritos com detalhes: 1) as atividades praticadas em sala de aula, que originaram as redações escolares estudadas neste estudo; 2) a metodologia de como foi realizado este estudo; 3) os contrastes realizados entre o *corpus* de redações das duas escolas e entre as redações, o texto de apoio e os textos dos livros didáticos; e os primeiros resultados e discussões.

### 5.1 Atividades geradoras

Os textos selecionados para essa análise inicial fizeram parte da unidade de estudos “Minha vida como filme”, realizada nas duas escolas, igualmente. Vale aqui salientar que a sistemática de avaliação é diferente em cada uma das escolas. Na escola estadual a avaliação segue uma escala numérica, enquanto na escola municipal é utilizado um sistema de conceitos descritivos.

Nessa unidade de estudo, os alunos leram e interpretaram o texto “A vida como filme” de Moacyr Scliar, publicado no jornal Zero Hora, no ano de 2003. O texto foi lido silenciosamente, para que os alunos tivessem um primeiro contato e, após, em voz alta, por alguns estudantes. Na sequência da leitura, algumas questões foram discutidas oralmente, como os gostos pessoais de filmes, atores e atrizes, além de como fariam um filme, caso fossem roteiristas do próprio enredo.

Após tais debates, os alunos realizaram atividades de interpretação de texto. Cada aluno foi estimulado a apresentar suas opiniões sobre o texto. Depois, no grupo, foram levantadas as possibilidades de interpretação e algumas questões que foram surgindo ao longo de cada discussão. Por fim, os educandos foram instigados a escreverem os seus textos, baseados nas leituras e nas interpretações discutidas. A proposta para a escrita foi a seguinte:

Quadro 1 – Proposta de produção textual.

Agora é a sua vez: escreva um texto contando como seria o filme de sua vida. Explique quais seriam os atores do filme, o que aconteceria na sua história, o que você destacaria e o que você eliminaria em seu roteiro.

Seja criativo e boa produção!

(mínimo: 25 linhas)

Fonte: acervo pessoal da professora (2018).

Os alunos tiveram dois períodos (cerca de 90 minutos) para realizar a tarefa, que compreendia escrever um texto relatando como seria o filme de sua vida, ou seja,

quais acontecimentos pessoais estariam nas telas do cinema e quais atores os representariam, caso pudessem controlar o próprio enredo.

As turmas, de maneira geral, pareceram bastante animadas com a proposta, pois demonstraram isso desde o início da unidade de estudos, na leitura do texto de apoio e nas discussões sobre filmes, atores e atrizes. Acreditamos que o assunto faz parte da realidade dos adolescentes, uma vez que o acesso a filmes e séries está cada vez mais facilitado, com a popularização de plataformas *on-line* como Netflix ou Now.

## 5.2 Metodologia do estudo piloto

Do total de produções entregues pelos alunos, 18 de cada turma atenderam a proposta como um todo (tanto em relação ao tema, quanto em relação ao número mínimo de linhas). Desse universo, retiramos quatro textos de cada turma para um exame detalhado. Essa amostra representa cerca de 20% do total produzido nessa unidade de ensino.

Os quatro textos de cada turma foram escolhidos, primeiramente, por atenderem satisfatoriamente à proposta de produção textual, e, depois, por serem semelhantes em quantidade na representação de cada escola e com avaliações (realizadas previamente pela professora) similares, conforme cada sistema escolar. A amostra constituída procurou alguma regularidade dos textos para exame, como o número de linhas de cada um. Assim, embora tenhamos um conjunto variado, tentamos obter uma amostra de dimensão mais ou menos equivalente em termo de tamanho do texto entre as duas escolas. Assim, a amostra do estudo está constituída conforme sintetiza o Quadro 2 a seguir:

Quadro 2 – textos por escola.

E.M.E.F. Décio Martins Costa	E.E.E.F. Dr. Ferreira de Abreu
---------------------------------	-----------------------------------

TURMAS C21 e C22	TURMA 81
4 textos	4 textos

Fonte: elaborado pelo autor (2019).

E, dentro dos quatro textos de cada turma, a seleção amostral tem a organização representada no Quadro 3 a seguir:

Quadro 3 – Relação por escola: texto, número de linhas e avaliação.

TEXTO (prefeitura)	LINHAS (mínimo:25)	AVALIAÇÃO (valor: A)	TEXTO (estado)	LINHAS (mínimo:25)	AVALIAÇÃO (valor:3,0)
1	31	A	1	27	3,0
2	38	AP	2	36	2,5
3	25	AP	3	27	2,0
4	25	AM	4	26	1,6

Fonte: elaborado pelo autor (2019).

No Quadro 3, observam-se as características específicas das redações dos alunos, como o número de linhas e a avaliação que cada texto recebeu. Vale esclarecer que o resultado das avaliações é demonstrado de maneira diferente nas duas redes escolares. Nas escolas da rede municipal, a avaliação é por conceitos e cada atividade avaliativa recebe um conceito que vai de NA (não atingiu os objetivos) até A (atingiu os objetivos).

Nas escolas estaduais, a avaliação é numérica, então, para essa atividade de produção textual foi estabelecido o valor de 3,0 pontos, no máximo, entre outras atividades avaliativas realizadas no trimestre. Para nosso estudo, foram selecionadas redações em que o número de linhas e a avaliação fossem semelhantes nos dois grupos, tentando equilibrar essas características entre os textos analisados.

Tais textos, conforme escritos, foram digitados, preservando-se sua grafia original e mantendo suas características na totalidade. A seguir, o Quadro 4 apresenta trechos das redações de cada grupo de alunos:

Quadro 4 – Trechos de redações de cada grupo de alunos.

GRUPO A (escola municipal)	GRUPO B (escola estadual)
Seria um filme bem grande, com momentos tristes e felizes, e os atores e atrizes seria Meus amigos e minha família. Eu contaria tudo, desde de quando eu nasci até os dias de hoje.	Meu Filme seria assim: num planeta muito distante existe uma raça chamada patoscidym, eles são uma raça misturada entre humanos e patos e o nome do planeta seria, planeta Quak.
Se um filme fosse inspirado em minha vida ou se eu pudesse fazer, Eu faria com que tudo de ruim que acontece na minha vida se tornasse algo bom ou só diminuir o que me deixa triste.	No meu filme eu era uma criança muito pobre que sonhava ser jogador de futebol. Atores: Cristiano Ronaldo, Eduardo, Messi e Neymar.
Meu filme iria ser de superação onde e minha família venceríamos todas dificuldades. Primeiro eu iria tentar virar jogador de futebol que é meu sonho de pequeno iria tirar toda minha familia das dificuldades que passamos todos os dias.	Mas vamos falar do filme, eu seria o principal é óbvio Eu seria protagonizado por Denzel Washington e na minha história aconteceria um um assassinato e eu desvendaria o caso com ajuda dos meus amigos Mauro, Matheus, Alexandre.
Bom, o filme da minha vida seria meio que so minha rotina. Por que de manhã è escola depois meio dia casa ai a tarde depende o que vou fazer as vezes durmo ou saio mais o filme seria de manhã escola meia dia casa almoço e saio com minha amiga para dar uma volta ai encontrariamos mais amigos nosso.	Se eu pudesse fazer um filme da minha, eu provavelmente deveria que modificar e inventar muita coisa, afinal eu sou ainda muito nova e minha começou a me “livrar” só agora, e mesmo assim eu ainda faço poucas coisa, até arrisco dizer que minha infância teve mais aventuras que a minha pré adolescência.

Fonte: acervo pessoal da professora (2018).

Estes grupos de quatro textos de cada escola foram reunidos em arquivo único e processados com o auxílio da ferramenta AntConc (ANTHONY'S, 2018). Esse *software*, de acesso livre, é um conjunto de ferramentas para análise de *corpus*, que produz listas de palavras e de seus contextos, sendo útil para análise de textos.

O texto de apoio à proposta, de Scliar (2003), também foi digitado e processado com a ferramenta AntConc. O mesmo foi feito com os textos dos livros didáticos, aos quais os alunos foram expostos antes de produzirem a redação solicitada. Dessa forma, o *corpus* geral do estudo piloto foi composto conforme resumido no Quadro 5:

Quadro 5 – Composição geral do *corpus* desse estudo piloto.

	Número de redações	Texto de apoio	Textos – livros didáticos

Grupo A (escola municipal)	4 textos	1: A vida como filme, de Moacyr Scliar	1: A cara da vida moderna, de Walcyr Carrasco
Grupo B (escola estadual)	4 textos	1: A vida como filme, de Moacyr Scliar	1: Um bom sujeito (fragmento), de Antônio Carlos Olivieri

Fonte: elaborado pelo autor (2019).

Os textos dos livros didáticos e o texto de apoio foram usados para averiguar em que medida os alunos ativariam suas memórias curtas e/ou longas ao escreverem os seus textos. O texto de Scliar (2003), cuja leitura foi a mais recente antes das produções, representa uma memória curta. Por outro lado, os textos dos livros didáticos, trabalhados algum tempo antes das produções, equivalem a uma memória longa. A ideia foi pesquisar o quanto e quais itens do vocabulário desses textos seriam utilizados pelos jovens nas suas redações, ou seja, em que medida influenciariam a sua escrita no que diz respeito ao perfil de vocabulário.

Desse material, nesse primeiro momento, com apoio da ferramenta AntConc, foram examinados:

- o número total de palavras (*tokens*);
- o número de palavras diferentes que são repetidas em um texto (*types*);
- a frequência de *types*;
- as diferenças entre as ocorrências de palavras lexicais e de palavras gramaticais;
- a riqueza lexical (relação entre *tokens* e *types*, que serve para medir a variedade do vocabulário de um texto);
- o universo de palavras de frequência um (*hapax legomena*).

A partir desses elementos, os dados foram levantados e contrastados, os quais são descritos a seguir.

### 5.3 Contrastes, resultados e discussões



Conforme já mencionado, as redações dos estudantes foram selecionadas de modo que obtivéssemos uma amostra homogênea e tão representativa quando possível. Desta amostra, o primeiro resultado obtido foi o número total de palavras do conjunto de quatro textos de cada grupo. São 836 palavras nos textos da escola municipal, uma média de 209 palavras por texto, frente a um universo de 729 palavras na escola do estadual, com média de 182 palavras por texto. Isso mostra uma aproximação entre os dois grupos, nesse aspecto.

Outro fator analisado foram as 20 primeiras palavras mais frequentes no *corpus* de estudo. Neste caso, também não foram obtidos resultados muito diferentes de uma escola para outra, conforme o Quadro 7 a seguir:

Quadro 6 – 20 palavras mais frequentes/textos dos alunos.

Textos da escola municipal			Textos da escola estadual		
Posição	Frequência	Palavra	Posição	Frequência	Palavra
1	33	e	1	33	e
2	27	que	2	28	eu
3	22	a	3	20	meu
4	19	de	4	18	a
5	19	não	5	16	minha
6	14	uma	6	16	que
7	13	as	7	15	filme
8	13	eu	8	15	seria
9	13	filme	9	14	no
10	13	mais	10	13	muito
11	12	minha	11	13	o
12	12	um	12	12	eles
13	11	o	13	12	para
14	11	vida	14	11	uma
15	10	os	15	8	me
16	10	para	16	8	se
17	10	pessoas	17	7	com
18	9	com	18	7	de
19	8	seria	19	7	pessoas
20	7	da	20	7	um

Legenda: Azul – Palavras lexicais em comum.

Fonte: elaborado pelo autor (2019).

No Quadro 6 é possível perceber o predomínio já esperado das palavras gramaticais em meio ao todo do vocabulário. Esse tipo de item corresponde à organização gramatical da língua. Nesse grupo, temos, entre os 20 primeiros itens mais frequentes, as conjunções (e, se), as preposições (de, da, para, com, no), os pronomes (ele, eles, eu, me, que, minha, meu, muito) e os artigos (a, as, um, uma, o, os). A primeira palavra, também gramatical, e que, inclusive, se repete em ambos os grupos, com a mesma quantidade de ocorrências é o E.

No grupo das palavras lexicais, que correspondem a unidades com conteúdo autônomo, como são substantivos, adjetivos e verbos, vemos que filme, pessoas e seria são as mais empregadas nos dois grupos. Esse grupo de palavras, que corresponde ao conjunto das primeiras palavras lexicais mais frequentes de um texto, tende a sinalizar os tópicos principais ou assuntos tratados nesse texto. Tais palavras também aparecem como as mais frequentes no texto de apoio. Entretanto, nenhuma delas aparece entre as 20 mais frequentes nos textos dos livros didáticos. Isso é o que busca ilustrar o Quadro 7 a seguir. Nesse quadro, vemos um levantamento das palavras mais frequentes nos textos que, conforme supusemos, serviriam como algum insumo para o vocabulário verificado nas redações:

Quadro 7 – 20 palavras mais frequentes – textos de contraste.

Texto: A vida como filme (memória curta)			Texto escola municipal: A cara da vida moderna (memória longa)			Texto escola estadual: Um bom sujeito (fragmento) (memória longa)		
Posição	Frequência	Palavra	Posição	Frequência	Palavra	Posição	Frequência	Palavra
1	15	filme	1	26	de	1	39	a
2	14	que	2	17	um	2	26	o
3	14	um	3	13	e	3	24	de
4	13	de	4	12	em	4	14	que
5	13	o	5	12	o	5	13	reinaldo
6	12	e	6	10	a	6	13	sujeito
7	9	uma	7	10	é	7	12	para
8	7	nosso	8	7	não	8	12	é
9	7	não	9	6	mais	9	11	não
10	6	a	10	6	se	10	11	professora
11	6	ao	11	5	com	11	10	do

12	6	do	12	5	mas	12	10	se
13	6	para	13	5	meu	13	8	um
14	5	em	14	5	os	14	7	da
15	5	os	15	5	para	15	7	e
16	5	se	16	5	que	16	7	eduardo
17	5	vida	17	5	uma	17	6	márcia
18	4	com	18	4	agora	18	6	no
19	4	história	19	4	no	19	6	núcleo
20	4	na	20	3	amigo	20	6	oração

Legenda: Azul – Palavras lexicais em comum com os textos dos alunos.

Roxo – Palavras lexicais que só aparece no texto em questão.

Fonte: elaborado pelo autor (2019).

A partir do Quadro 7, em relação aos textos dos livros didáticos, tomados como parâmetro de memória longa do redator, é possível inferir que os alunos não utilizaram palavras lexicais desse universo. Desses textos, foram semelhantes somente alguns itens gramaticais, que também são os de maior frequência na língua portuguesa como um todo. Essa, então, não parece uma escolha influenciada pelos textos dos livros didáticos. Por outro lado, os alunos parecem ter buscado, ao escrever seus textos, muitos itens do vocabulário do texto de apoio. Ativaram, assim, em tese, mais as suas memórias curtas do que as longas.

Algo que também provavelmente aconteceu foi o resgate das palavras lexicais a partir da proposta de produção textual, visto que ela instigava o aluno a escrever sobre como seria o filme da sua vida, caso pudessem ser roteiristas do próprio enredo. Portanto, as palavras filme, vida e seria funcionam como tópicos prévios à produção do texto, com alta frequência esperada.

A riqueza lexical é uma medida tradicional em estatística linguística. Ela é entendida como “[...] uma relação que se estabelece entre número total de palavras repetidas e diferentes de um texto e o número total de palavras nele encontradas.” (SILVA, 2011, p.36). Essa medida pode ser representada pela seguinte fórmula: Riqueza Lexical = número de *types* X 100 dividido pelo número de *tokens*. Segundo Berber Sardinha

(2004), quanto maior for o número de *types*, maior também será a variedade do vocabulário, já que este terá mais palavras diferentes e, da mesma forma, quanto menor for este número, menor será a variedade, ou seja, o texto analisado é mais repetitivo.

Essa proporção, que mostra o quanto um vocabulário é mais ou menos variado em um dado universo de texto ou *corpus*, praticamente não variou entre os textos de uma escola para outra. Isso é o que ilustra o Quadro 8 a seguir:

Quadro 8 – Riqueza Lexical.

	Riqueza Lexical
Escola Municipal	41,8%
Escola Estadual	41,9%
Texto de Apoio	52,3%
Texto Didático (Município)	61%
Texto Didático (Estado)	40,8%

Fonte: elaborado pelo autor (2019).

A partir destes dados, foi possível perceber que a variação de vocabulário empregado nas redações dos dois grupos foi praticamente idêntica (41,8% e 41,9%).

Em relação à diversidade vocabular, um artigo científico na área de Biologia apresenta cerca de 29% de riqueza lexical (FINATTO; CREMONESE; AZEREDO, 2008). Nesse sentido, tais materiais repetem propositadamente os termos e as palavras específicas da área como forma de reiterar o assunto tratado, são, portanto, considerados como referência de repetitividade.

Por outro lado, um texto jornalístico tende a seguir a regra da boa redação jornalística de não repetir palavras. Nesse grupo se insere um artigo jornalístico premiado na temática do meio ambiente, que apresenta 40% de riqueza lexical (BISOGNIN, 2009). Em função disso, os textos examinados no estudo piloto aproximam-se da riqueza de um texto jornalístico e, por isso, podem ser considerados com vocabulário escrito variado.

Os textos dos livros didáticos, conforme observado no Quadro 6, distanciam-se desses padrões de comparação apresentados. Em função disso, será preciso reavaliar a presença dos textos dos livros didáticos como contraste.

Para que se possa realmente analisar a variedade do vocabulário dos textos escolares, foi importante considerar o universo de palavras de frequência 1 (*hapax legomena*) em cada grupo. Afinal, se considerarmos faixas de frequências mais elevadas, temos o perfil gramatical da língua em ação. Do mesmo modo, a faixa dos itens lexicais de maior frequência em um texto tende a corresponder aos temas que serão tratados.

Nesse segmento de palavras de frequência única, vemos a particularidade do vocabulário de um texto ou mesmo de um dado redator. Nessa amostra inicial de redações, verificamos o padrão de *hapax legomena*, sintetizado no Quadro 9, a seguir. Esse quadro retrata o universo vocabular representado pelo recorte das palavras que iniciam com a letra P e que são empregadas apenas uma vez.

Quadro 9 – Amostra de *hapax legomena* dos dois grupos de redações escolares.

	GRUPO A (escola municipal)	GRUPO B (escola estadual)
1.	palhaçada	paveva
2.	parecidos	pagaram
3.	participar	parte
4.	paz	partes
5.	pela	partiria
6.	pelos	passa
7.	perfeita	pegaria
8.	peessoa	pensei
9.	piscina	perdido
10.	possiveis	perigosas
11.	praça	preferido
12.	preça	<b>primeira</b>
13.	prima	<b>primeiro</b>
14.	<b>primeiramente</b>	professoras
15.	<b>primeiro</b>	proposta
16.	pro	proternizado
17.	<b>pudecem</b>	<b>pudesse</b>
18.	<b>pudese</b>	
19.	público	

Legenda: Negrito = palavras comuns entre os dois grupos. Fonte: elaborado pelo autor (2019).

Conforme se observa, o número de itens nesse segmento da letra P é bastante aproximado entre os dois grupos: 19 palavras no GRUPO A frente a 17 palavras no GRUPO B. Entretanto, nessa pequena amostra, o percentual de coincidência de vocabulário, considerando uma base semântica de palavras, conforme assinalado em **negrito** no Quadro 9, é de cerca de 19,5% somente. Isto é, há uma grande diferenciação entre os dois grupos, cada um prefere utilizar palavras distintas. Apenas as palavras: primeiramente, primeiro, pudecem, pudese, primeira, primeiro e pudesse foram consideradas semanticamente semelhantes.

A partir disso, é possível acreditar que a análise dos itens *hapax legomena* pode ser bastante promissora em relação à pesquisa como um todo.

A seguir, retomamos nossas hipóteses iniciais, agora de acordo com o que foi percebido na primeira análise de pequena amostra do nosso *corpus* de estudo.

## 6 Retomada das hipóteses

Observou-se que, apesar da quantidade total de palavras ser maior na escola da prefeitura, isso não interferiu na riqueza lexical, que se manteve em cerca de 42% nas duas escolas. Com base nestes dados, é possível afirmar, então, que os alunos não foram repetitivos em relação às palavras utilizadas, pois a variedade vocabular dos alunos aproxima-se do padrão de variedade vocabular exibido por um texto jornalístico (40%). Em síntese, não se confirma o que imaginamos conforme nossa experiência docente.

Quanto à segunda hipótese colocada, segundo a qual o vocabulário dos estudantes tenderia a reproduzir o dos professores e/ou da proposta da redação e dos materiais de apoio, verificou-se a sua confirmação. Isso pode ser atestado pela observação de coincidência entre o vocabulário mais frequente nas orientações de escrita, no texto de apoio, bem como nas redações dos alunos. O tema principal da unidade de ensino

(filme) e as palavras que o rodeiam (vida, seria), por exemplo, coincidem com as mais utilizadas pelos educandos nos seus textos.

Por outro lado, quanto à terceira hipótese, de acordo com as vinte primeiras palavras mais frequentes em ambos os acervos, foi possível perceber que não houve variação de uma escola para outra no que se refere ao desempenho vocabular. Afinal, nos dois grupos, as palavras que mais apareceram foram as gramaticais (o que é bem comum, na língua portuguesa como um todo), seguidas de filme, vida e seria, palavras lexicais presentes não só no texto de apoio, mas também na proposta de produção textual exposta aos alunos. Neste sentido, nossa última hipótese foi refutada, uma vez que o desempenho dos alunos da escola municipal, neste pequeno conjunto de textos analisados, não foi diferente dos estudantes da escola estadual. Em síntese, constatamos que ambos os públicos apresentaram desempenhos semelhantes.

## 7 Avaliação e perspectivas

Este artigo relata uma exploração inicial de um conjunto de textos produzidos por alunos de duas escolas públicas do RS: uma municipal e outra estadual. Também foram considerados, no estudo, textos dos livros didáticos de Português utilizados nessas escolas e os textos de apoio para produção textual sob exame.

Todos os materiais de análise foram processados com a ferramenta AntConc para observação do perfil de vocabulário em termos de sua distribuição quantitativa. Inicialmente, foram analisados os dados: número de palavras diferentes (*types*) e o número total de palavras (*tokens*), o percentual de riqueza lexical, o léxico mais frequente, além dos *hapax legomena* exemplificados pelos itens que iniciam pela letra P.

Em síntese, das três hipóteses inicialmente colocadas, foi confirmada apenas uma, que mostra uma semelhança entre o vocabulário dos estudantes e dos textos de apoio.

Como o estudo é inicial, ainda há uma série de aspectos a serem considerados e repensados para uma observação em maior escala. Entre esses aspectos, ressaltamos os seguintes:

- avaliar a quantidade de textos necessária para estudo em grande escala. Para o estudo-piloto apresentado, todos os textos analisados eram de alunos do 8º ano das duas escolas. Essa é a homogeneidade mais expressiva no que diz respeito ao todo dos textos com que contamos para a pesquisa;
- ponderar o tratamento da irregularidade no que se refere ao tema desenvolvido em cada grupo de redações, pois há coincidência de tema em apenas 18 textos de cada escola. As outras redações referem-se a temáticas diferentes;
- ampliar o *corpus* de estudo, acrescentando cerca de 60 textos, todos escritos por alunos do 6º ano das duas escolas, sobre a mesma temática;
- reavaliar o papel do texto do livro didático como contraste ao *corpus* de estudo, uma vez que, conforme demonstrado no estudo, esses textos parecem não interferir no vocabulário escrito dos alunos.

Ao comparar o universo vocabular das redações dos dois grupos no que tange à quantidade total de palavras, à quantidade de palavras diferentes, às palavras mais frequentes (lexicais e gramaticais), à riqueza lexical e ao universo de palavras de frequência única, esse perfil de vocabulário escrito estará traçado.

Em termos do rendimento da análise preliminar, via estudo piloto, a camada de itens do tipo *hapax legomena* parece ser a mais promissora em termos de comparações. Isto é, os itens de frequência única empregados pelos estudantes, no seu todo, tendem a espelhar as escolhas mais particulares e a variedade do vocabulário dos textos.

A partir do que examinamos em uma pequena amostra e do que observaremos em uma grande amostra ampliada, objetivamos construir conhecimentos úteis sobre perfis de vocabulário escrito de estudantes do Ensino Fundamental. A pesquisa com



dados serve fundamentalmente para validar ou refutar impressões iniciais ou mesmo aquelas construídas ao longo de anos de atuação. Assim, queremos contribuir para a qualificação do trabalho do professor em sala de aula. Além disso, é importante compartilhar essa experiência com os colegas, tanto professores de língua materna, quanto pesquisadores da área das Ciências do Léxico.

### Referências bibliográficas

ANTHONY, L. (2019). **AntConc** (Versão 3.5.8) [Software de Computador]. Tóquio, Japão: Universidade de Waseda. Disponível em: <https://www.laurenceanthony.net/software>. Acesso em: 23 ago. 2019.

BASILIO, M. **Formação e classes de palavras no português do Brasil**. São Paulo: Contexto, 2004.

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri-SP: Manole, 2004.

BIDERMAN, M. T. **Teoria Linguística (lingüística quantitativa e computacional)**. Rio de Janeiro: Livros Técnicos e Científicos, 1978.

BIDERMAN, M. T. A face quantitativa da linguagem: um dicionário de frequências do português. **Alfa**, São Paulo, v. 42, número especial, p. 161-181, 1998.

BISOGNIN, T. R. **Sem medo do Internetês**. Porto Alegre: AGE, 2009.

CEREJA, W. R.; MAGALHÃES, T. C. **Português – Linguagens**. 8º ano do Ensino Fundamental, 8ed. São Paulo: Atual, 2014.

ENDRUWEIT, M. L. **A redação nota dez**. 2000. 189f. Dissertação de Mestrado (Mestrado em Estudos da Linguagem) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2000.

FINATTO, M. J. B. **Padrões do português popular escrito: o vocabulário do jornal diário gaúcho**. Fase 1. Projeto de Pesquisa contemplado Edital MCT/CNPq, nº14/2009 – Universal. Novembro de 2009. Disponível em: <http://www.ufrgs.br/textecc/porlexbras/porpopular/arquivos/PorPopularPropostaCNPqaprovada.pdf>. Acesso em: 05 abr. 2018.

FINATTO, M. J. B.; CREMONESE, L. E.; AZEREDO, S. O vocabulário na redação de vestibular: do enfoque estatístico às especificidades da enunciação. *In: ABREU, S. (org.) A redação no vestibular: do leitor ao produtor de texto.* COPERSE/UFRGS. Porto Alegre: Editora da UFRGS, 2008, p.95-108.

LUIZ, A. P. A. T. **Estudo da Redação Enem:** questões e perspectivas de uma práxis interdisciplinar e colaborativa. 2018. 171 f. Dissertação (Mestrado em Estudos Linguísticos) - Universidade Federal de Uberlândia, Uberlândia, 2018.

NUNES, P. A. Proficiência ou rendimento: o que a redação do vestibular avalia? *In: REBELLO, L. S.; FLORES, V. D. N. (org.) O Texto de Vestibular em Perspectiva.* COPERSE/UFRGS. Porto Alegre: Editora da UFRGS, 2014, p.53-67.

OLIVEIRA, T. A.; SILVA, E. G. O.; SILVA, C. O.; ARAÚJO, L. A. M. **Tecendo Linguagens – Língua Portuguesa.** 8º ano, 4ed. São Paulo: IBEP, 2015.

SILVA, B. R.; FINATTO, M. J. B. Português popular escrito: o vocabulário do jornal diário gaúcho. *In: X Salão de Iniciação Científica da PUCRS, 2009, Porto Alegre. Anais do X Salão de Iniciação Científica da PUCRS.* Porto Alegre: EDIPUCRS, 2009. p. 3332-3334.

SILVA, B. R. **Ensino de vocabulário no Ensino de Jovens e Adultos (EJA):** atividades a partir do *corpus* do jornal Diário Gaúcho. 2011. 83f. Trabalho de Conclusão de Curso (Graduação em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.

Artigo recebido em: 13.12.2019

Artigo aprovado em: 06.02.2020