

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

BRUNA RODRIGUES DA SILVA

**VOCABULÁRIO ESCRITO DE ESTUDANTES DE ESCOLAS PÚBLICAS DO RIO  
GRANDE DO SUL: UM ESTUDO LÉXICO-ESTATÍSTICO**

PORTO ALEGRE

2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
INSTITUTO DE LETRAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

BRUNA RODRIGUES DA SILVA

**VOCABULÁRIO ESCRITO DE ESTUDANTES DE ESCOLAS PÚBLICAS DO RIO  
GRANDE DO SUL: UM ESTUDO LÉXICO-ESTATÍSTICO**

Dissertação de Mestrado em Lexicografia, Terminologia e Tradução: Relações Textuais apresentada como requisito parcial para obtenção do título de Mestre em Estudos da Linguagem pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Maria José Bocorny Finatto

PORTO ALEGRE

2021

## FICHA CATALOGRÁFICA

### CIP - Catalogação na Publicação

Silva, Bruna Rodrigues da  
Vocabulário escrito de estudantes de escolas  
públicas do Rio Grande do Sul: um estudo  
léxico-estatístico / Bruna Rodrigues da Silva. --  
2021.

321 f.

Orientadora: Maria José Bocorny Finatto.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Instituto de Letras, Programa de  
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2021.

1. Redação Escolar. 2. Estatística Lexical. 3.  
Linguística de Corpus. 4. Língua Portuguesa. 5. Ensino  
Fundamental. I. Finatto, Maria José Bocorny, orient.  
II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os  
dados fornecidos pelo(a) autor(a).

BRUNA RODRIGUES DA SILVA

**VOCABULÁRIO ESCRITO DE ESTUDANTES DE ESCOLAS PÚBLICAS DO RIO  
GRANDE DO SUL: UM ESTUDO LÉXICO-ESTATÍSTICO**

Dissertação de Mestrado em Lexicografia, Terminologia e Tradução: Relações Textuais apresentada como requisito parcial para obtenção do título de Mestre em Estudos da Linguagem pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Aprovada em 25 de fevereiro de 2021.

Resultado: Aprovado, conceito A.

BANCA EXAMINADORA

---

Bianca Franco Pasqualini  
Instituto de Letras  
Universidade Federal do Rio Grande do Sul (UFRGS)

---

Eduardo Batista da Silva  
Instituto de Letras  
Universidade Estadual de Goiás (UEG)

---

Guilherme Fromm  
Instituto de Letras e Linguística  
Universidade Federal de Uberlândia (UFU)

*Aos meus alunos, minha fonte diária de inspiração.*

## AGRADECIMENTOS

Aos colegas do grupo de pesquisa em Acessibilidade Textual e Terminológica (ATT), pela parceria e pela troca de conhecimentos, em especial, ao Lucas Tcacenco e ao Gabriel Ponomarenko. Ao Lucas pelos *abstracts*, por ser minha dupla desde o início dos estudos no PPG e por estarmos sempre juntos, dividindo as angústias, as dúvidas, as leituras, as discussões, os alunos, os estudos piloto e os artigos científicos. E ao Gabriel pela árdua tarefa de digitação de todo *corpus* de estudo, além do incansável suporte (praticamente on-line) em todos os questionamentos de informática, programação, *software* e *hardware*.

À colega de pesquisa de longa data, professora titular da cadeira de Léxico e Dicionários em que realizei meu estágio de docência, Bianca Pasqualini. Obrigada pela disponibilidade, pelas discussões e pelas contribuições ao longo do período de estágio. Aos alunos da disciplina de Léxico e Dicionários 2019/2 da Graduação em Letras da UFRGS: muito obrigada! Vocês não imaginam o quanto contribuíram para este estudo.

À outra colega de pesquisa também da mesma longa data, Aline Evers, pela leitura atenta deste trabalho, pelas considerações e por toda ajuda na reta final.

Ao meu eterno diretor preferido, Paulo Ricardo de Oliveira Fraga, por estar sempre disponível, por tentar encontrar solução para tudo, por nunca me dizer não. A minha (também eterna) vice, Ana Paula Taday, por encaminhar com rapidez e eficiência todos os meus processos SEI (que não foram poucos), por me emprestar seus dicionários para prova de proficiência e por prontamente atualizar meu ponto (que eu sei bem que também não foi tarefa fácil). A minha super, Lilian Rocca, por ajustar os horários da escola em tempo recorde, sempre que necessário, e por ser um exemplo de profissionalismo e de eficiência. À colega Camila Bressan, por gentilmente contribuir com as redações dos seus alunos de B30, para que fosse possível o contraste entre os anos/ciclos. À prefeitura municipal de Porto Alegre, por me conceder licenças para estudos e para eventos sempre que solicitadas. Nesse sentido, como é bom ser servidor municipal.

A minha outra vice, Ione Monteiro, por modificar os horários da escola conforme o que eu precisava para assistir às aulas do mestrado. À colega, amiga e parceira Andreia Ferreira de Souza, por ouvir minhas lamentações, reclamações e crises existenciais ao longo de todo esse processo. Ao governo do estado do Rio

Grande do Sul, por me conceder uma Licença Qualificação Profissional em tempos tão difíceis. E por, logo em seguida, me conceder Licença Interesse Particular, para que fosse possível dedicação aos estudos de Pós-Graduação.

Aos meus colegas de magistério, que sempre me incentivaram e me apoiaram ao longo desses anos, por cada palavra amiga, abraço, apoio, interesse e disponibilidade com o meu estudo.

Ao PPG-Letras/UFRGS por tudo. Por todos os ensinamentos, pelo crescimento, pela disponibilidade, pela prontidão e por entender o momento difícil pelo qual todos passamos ao longo do ano de 2020. Ao Instituto de Letras e à UFRGS, por me abrigar mais uma vez e por fazer eu me sentir em casa novamente, mesmo depois de tantos anos. Ao Canísio, melhor funcionário do PPG, por me acalmar, por ajudar a me organizar e por sempre ter uma resposta e um ombro amigos para todos os meus pequenos surtos durante esses mais de dois anos.

À CAPES, por conceder bolsa de estudos para professores de educação básica da rede pública, e por continuar investindo na formação desses profissionais mesmo depois de formados e atuantes nas escolas públicas. Por prorrogar seus prazos e auxílios em função da pandemia, contribuindo mais uma vez para que as pesquisas fossem concluídas com a paz que o momento exigiu.

Ao pessoal do NILC-USP, sobretudo ao prof. Dr. Thiago Pardo e seu orientando Angelo Tessaro, pela grande ajuda e disponibilidade de lematização do *corpus*, sem a qual essa pesquisa não seria a mesma. Ao também integrante do NILC, Nathan Siegle Hartmann, por prontamente dividir os resultados da sua tese, contribuindo com os resultados da minha dissertação. À revista GTLex, pela publicação dos artigos. Ao prof. Dr. Guilherme Fromm, pela geração das *keywords* e pela disponibilidade de sempre. Ao prof. Dr. Tony Berber Sardinha, pela gentileza de nos conceder dados do *corpus* Banco de Português que não estavam mais disponíveis on-line. Como diria minha orientadora: quem tem amigos, tem tudo!

Aos professores que aceitaram participar das considerações e da avaliação deste trabalho: muito obrigada!

A minha orientadora, parceira e amiga de longa data, Maria José Bocorny Finatto, sem a qual eu não teria retornado aos estudos e à pesquisa acadêmica. Por acreditar e confiar em mim e no meu trabalho, concomitante à sala de aula; por ouvir minhas lamúrias e reclamações, sempre encontrando um meio de acreditar e de

fazer as coisas darem certo. Muito obrigada por tanto. Te admiro muito mais do que tu imaginas. Quando eu crescer, quero ser igual a ti!

Aos meus amigos, pois é impossível ser feliz sozinho. Por me aguentarem, por entenderem os meus sumiços e as várias recusas durante esse período, por saberem e por me lembrarem que eu estava no caminho certo, por torcerem e por vibrarem comigo a cada novidade, a cada descoberta e cada nova conquista (que não foram poucas). Muito obrigada! Em especial ao Eduardo Belmonte de Souza, à Paula Cauduro Bianco e à Vanessa Gaitkoski. Ao Eduardo, inicialmente colega de faculdade e de profissão e hoje colega de Pós-Graduação, por dividir comigo os anseios da vida docente e, agora, docente e acadêmica; e a mudança do retorno à sala de aula como alunos. À Paula, por me ouvir incondicionalmente, por torcer como se fosse da família e por comemorar as minhas vitórias como se fossem dela, por ser o meu Coração. E a amiga de longa data, desde a barriga das nossas mães, Vanessa Gaitkoski, por às 22h30min de uma sexta-feira, no meio de um voo internacional, me ajudar a fazer contas, proporções e ainda oferecer gráficos caso necessário (mal sabia ela o quanto – e quantos – seriam necessários depois disso); pela disponibilidade, pela ajuda e pela amizade. Mais uma vez: quem tem amigos, tem tudo!

Aos meus dindos, Gilberto e Luiza Polônio, por estarem sempre ao meu lado e por se fazerem presentes a minha vida toda, cumprindo com excelência o papel de padrinhos assumido há mais de 30 anos. Iguamente à Camila e à Luiza Sommer, sempre presentes e sempre torcendo, independente da distância física ou em quilômetros, por me mostrarem que as verdadeiras amizades crescem mesmo a longas distâncias.

Ao Afonso e a Marion Zampieri, por me incluírem nas suas orações e por torcerem pelo meu sucesso pessoal e profissional.

À *golden* Luna Freya, por chegar de mansinho e me fazer aprender tanto, sobretudo sobre o amor de quatro patas, cheio de pelos e de baba. À pequena Gigi, por me mostrar como é bom ter um rabinho abanando quando acordamos ou chegamos depois de um dia exaustivo.

Ao Guilherme Morais Zampieri, pelos gráficos, imagens e tabelas; pelo passador de slides que transformou as minhas apresentações acadêmicas; pela cadeira que pôs minha coluna de volta no lugar; por entender minhas ausências, minhas dores de cabeça, minhas reclamações, meu cansaço, minhas leituras,



minhas reuniões e meus compromissos; por me incentivar desde sempre; por caminhar junto comigo e por fazer planos; pela felicidade em forma de amor.

A minha família, meu amor maior. A minha vó Ceni, por ser minha companheira, por me tornar a neta preferida, pelo amor em forma de pão e pizza feitos à mão. A minha tia Simone, pela parceria de sempre, por ser de tudo um pouco para toda nossa família desde sempre. A meu tio Flávio, pelo primeiro computador, quando eu passei na UFRGS em 2007, incentivando meus estudos e minha vida acadêmica. A minha prima Diandra, por lá em 2006, quando eu não passei no primeiro vestibular, me ensinar que: o que é meu, de mim ninguém tira! A minha irmã Bianca, por depois de 10 anos aceitar dividir tudo que ela tinha só para ela comigo (inclusive a profissão). Ao Anderson, por desvendar o mistério do meu notebook e por ajustá-lo numa fração de segundos. A minha mãe, por todos os almoços prontos na hora; pela casa, comida e roupa lavada; por me aguentar em casa o dia inteiro na frente de um computador. Ao meu pai, por acordar todas as manhãs para me levar na parada de ônibus durante anos, inverno e verão, na chuva ou no sol; por cuidar de mim, do meu quarto, do meu carro, dos meus documentos e de tudo mais que for preciso; por fazer tudo por mim e por nós desde sempre. A minha avó Marina Rosa da Silva que, com certeza, está radiante e orgulhosa lá no céu (*in memoriam*). Eu AMO vocês!

Muito obrigada!

Sem vocês, eu nada seria.

*Se é verdade que não  
existe língua sem gramática, mais  
verdade ainda é que sem léxico  
não há língua. As palavras são a  
matéria-prima com que  
construímos nossas ações de  
linguagem.*  
(ANTUNES, 2014, p. 27)

## RESUMO

Este trabalho descreve resultados de uma pesquisa de mestrado inserida no âmbito dos estudos lexicológicos de enfoque estatístico, baseados em *corpora*. O objetivo principal é a descrição e a análise contrastiva do perfil de vocabulário escrito empregado por estudantes dos anos finais do Ensino Fundamental de duas escolas públicas do estado do Rio Grande do Sul – Brasil. O *corpus* de estudo é composto por um conjunto de 154 redações desidentificadas de alunos de 6º ano e de 8º ano de duas escolas, uma administrada pelo Governo do Estado e outra pelo Município de Porto Alegre. Ambas as escolas se localizam na Zona Norte da capital gaúcha, próximas entre si e na fronteira da região metropolitana. O *corpus* de estudo é heterogêneo quanto às propostas geradoras das redações, mas homogêneo quanto à quantidade de textos por escola e por ano escolar. O *corpus* foi examinado em dois grupos principais, de acordo com a procedência por escola. Parte-se da ideia de que o perfil vocabular seria diferente entre os dois grupos de textos, em função das diferentes condições de infraestrutura em cada uma das escolas e de relatos de alguns professores. Como contraste para o exame do seu vocabulário, são usados: os textos de apoio para as produções textuais – quando existentes – e as propostas que orientaram a escrita dos alunos. Como *corpora* de referência para um Português escrito fundamental ou básico, conforme utilizados por pessoas adultas, são utilizados: o CorPop, o Banco de Português, o Corpus Brasileiro e o Lácio Web. Os três últimos foram utilizados como referência somente para identificação de palavras-chave das redações do *corpus* de estudo. O CorPop, por sua vez, forneceu uma referência sobre simplicidade do vocabulário empregado. Com apoio de ferramentas computacionais, da estatística linguística e da Linguística de Corpus, as redações foram descritas, analisadas e comparadas entre si e com os materiais de contraste e com os *corpora* de referência. Os resultados indicam que a riqueza vocabular exibida nas redações não tende a variar entre as produções dos estudantes da escola municipal (16,5%) e as dos estudantes da escola estadual (16,5%). Do mesmo modo, não há muita diversidade de palavras empregadas entre os dois grupos quando se considera as mais frequentes de uma mesma proposta de tema. O grupo de palavras com frequência única – *hapax legomenon* – empregado nas redações mostrou-se como o principal fator de especificidade do vocabulário de cada grupo. Com relação ao aspecto de simplicidade do vocabulário, verificou-se

uma coincidência de quase 70% com o CorPop, ainda que esse percentual possa ser ajustado para cobrir formas não lematizadas das palavras e itens não contemplados nesse *corpus*. Por fim, o trabalho refuta a ideia de que os dois grupos possuiriam perfis de vocabulário escrito diferentes e apresenta algumas sugestões de atividades para professores do Ensino Fundamental tratarem do tema do vocabulário em suas aulas.

**Palavras-chave:** Redação escolar. Vocabulário. Estatística lexical. Lexicologia. Linguística de Corpus. Ensino Fundamental. Língua Portuguesa.

## ABSTRACT

The aim of this thesis is to describe and present the findings of an investigation in the area of Lexicology using a statistical approach based on corpora. The main objective is the description and contrastive analysis of the written vocabulary of Middle School students from two public schools of the state of Rio Grande do Sul – Brazil. The study corpus comprises 154 writing samples of 6<sup>th</sup> and 8<sup>th</sup> year students from two public schools: one run by the Government of the State, and another run by the Municipal Government of Porto Alegre. Students' identities have been concealed in all writings. Both schools are located in the Sarandi area, in the north side of Porto Alegre. The study corpus is heterogeneous in terms of the topics presented to the students, but homogeneous in terms of the number of writing samples per school and per school year. The corpus was divided into two main groups, one group per school. It was assumed that the vocabulary to be used would be different between the two groups because of the different infrastructure conditions in each school and due to reports from the teachers. The following items were used to contrast and analyze the vocabulary: the issue topic for the written production – whenever applicable – and the rubrics for the activity. As a reference for elementary-level Portuguese used by adults, the following corpora were selected: CorPop, Banco de Português, Corpus Brasileiro and Lácio Web. The last three corpora were used as reference for identifying keywords in the writing samples of the study corpus and CorPop provided a reference on the simplicity of the vocabulary used. The description, analysis and comparisons of the written productions with one another and against the contrast corpora were aided by computer tools, Statistical Linguistics and Corpus Linguistics. The data from the study corpus and reference corpus have been compared and contrasted. Results have shown that vocabulary richness in the students' texts tends to show little to no variation in both schools (16.5%). Similarly, there is no diversity of words between the two groups when looking for the most frequent ones in a single topic. The only difference between the productions of both groups was on words or expressions used that occurred only once – *hapax legomenon* – in the writing samples. Regarding the level of simplicity of the vocabulary, there was a coincidence of almost 70% with CorPop vocabulary, although this percentage can be adjusted to cover non-lemmatized forms of words and items not included in this corpus. Therefore, this investigation refutes the claim that both groups would have different

vocabularies. Lastly, suggestions of activities for Middle School teachers to work with vocabulary under a linguistic analysis perspective in class are presented based on the findings of this study.

**Keywords:** School writing. Vocabulary. Lexical Statistics. Lexicology. Corpus Linguistics. Middle School. Portuguese Language.

## LISTA DE ABREVIATURAS E SIGLAS

BNCC	Base Nacional Comum Curricular
Celpe-Bras	Certificado de Proficiência em Língua Portuguesa para Estrangeiros
CV-UFRGS	Concurso Vestibular da Universidade Federal do Rio Grande do Sul
Enem	Exame Nacional do Ensino Médio
INAF	Indicador de Analfabetismo Funcional
LAEL	Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem
LC	Linguística de Corpus
NILC-USP	Núcleo Interinstitucional de Linguística Computacional
PCN	Parâmetros Curriculares Nacionais
PLN	Processamento de Linguagem Natural
PNLD	Programa Nacional do Livro e do Material Didático
RCs	Referenciais Curriculares
UFRGS	Universidade Federal do Rio Grande do Sul
USP	Universidade de São Paulo

## LISTA DE FIGURAS

Figura 1 – Trecho de texto original manuscrito I .....	28
Figura 2 – <i>Keywords</i> – modelo .....	75
Figura 3 – Trecho de texto original manuscrito II .....	112
Figura 4 – Dados do CorPop .....	118
Figura 5 – Palavras frequentes em relação aos textos de apoio .....	145
Figura 6 – Riqueza Lexical – textos de apoio .....	147
Figura 7 – Informações gerais – estudos piloto e experimento principal .....	172
Figura 8 – Riqueza Lexical – estudos piloto e experimento principal .....	173
Figura 9 – Porcentagem de <i>hapax</i> coincidente – estudos piloto e experimento principal .....	174
Figura 10 – Nuvem de palavras gramaticais .....	191
Figura 11 – Nuvem de palavras lexicais .....	193
Figura 12 – <i>Hapax</i> em comum – Letra A .....	196
Figura 13 – <i>Hapax</i> em comum – Letra P .....	197
Figura 14 – Nuvem de palavras – Texto de apoio / <i>corpus</i> de estudo .....	200
Figura 15 – Síntese das hipóteses .....	208



## LISTA DE GRÁFICOS

Gráfico 1 – Gráfico de frequência de palavras .....	65
Gráfico 2 – Camadas de um gráfico de frequência de palavras.....	121
Gráfico 3 – Gráfico de frequência de palavras – Estudo Piloto IV.....	157
Gráfico 4 – <i>Hapax</i> – Estudo Piloto IV .....	158
Gráfico 5 – <i>Hapax</i> – Estudo Piloto IV (em ponderação).....	158
Gráfico 6 – Gráfico de frequência de palavras – Experimento Principal .....	163
Gráfico 7 – <i>Hapax</i> – Experimento Principal.....	164

## LISTA DE QUADROS

Quadro 1 – Trecho de redação/ trecho de texto de apoio .....	26
Quadro 2 – Trecho de texto original digitado I.....	28
Quadro 3 – Critérios para formação de um <i>corpus</i> .....	54
Quadro 4 – Termos e explicações.....	62
Quadro 5 – Trecho de texto original digitado II.....	112
Quadro 6 – Composição e características do <i>corpus</i> de estudo – amostra sob exame .....	114
Quadro 7 – Classificação do <i>corpus</i> .....	115
Quadro 8 – Temáticas do <i>corpus</i> de estudo.....	116
Quadro 9 – Ano/ciclo do <i>corpus</i> de estudo.....	116
Quadro 10 – Escolas do <i>corpus</i> de estudo.....	116
Quadro 11 – Textos de apoio para os redatores – usados como contraste .....	117
Quadro 12 – Textos didáticos utilizados como contraste .....	117
Quadro 13 – Textos do primeiro estudo piloto.....	126
Quadro 14 – Composição geral do <i>corpus</i> do estudo piloto I.....	127
Quadro 15 – Composição geral do <i>corpus</i> do estudo piloto II.....	136
Quadro 16 – Palavras lexicais e gramaticais em relação ao texto de apoio – estudo piloto II.....	138
Quadro 17 – <i>Corpus</i> de estudo – estudo piloto III .....	141
Quadro 18 – Textos de contraste – estudo piloto III .....	141
Quadro 19 – <i>Hapax</i> – estudo piloto III .....	148
Quadro 20 – <i>Hapax</i> lematizado – estudo piloto III.....	149
Quadro 21 – <i>Corpus</i> do estudo piloto IV .....	151
Quadro 22 – 50 palavras mais frequentes 6º/8º ano .....	153
Quadro 23 – <i>Hapax</i> – 6º e 8º anos.....	155
Quadro 24 – Composição geral do <i>corpus</i> – experimento principal .....	160
Quadro 25 – Proporção e frequência /palavras gramaticais e lexicais – experimento principal.....	161
Quadro 26 – <i>Hapax</i> legomena – experimento principal.....	165
Quadro 27 – Palavras semanticamente comuns entre o <i>corpus</i> de estudo e o CorPop .....	166

Quadro 28 – Lista de palavras-chave do <i>corpus</i> de estudo em relação aos <i>corpora</i> de referência .....	169
Quadro 29 – Exemplos de contextos das palavras gramaticais mais usadas .....	177
Quadro 30 – Exemplos de contextos das palavras lexicais mais usadas .....	178
Quadro 31 – Exemplos de contextos de palavras-chave .....	181
Quadro 32 – Composição dos grupos de alunos pesquisados .....	183
Quadro 33 – Exemplos de uso do “porque” – Grupo A .....	183
Quadro 34 – Exemplos de uso de palavras lexicais – Grupo A.....	184
Quadro 35 – Exemplos de uso do “como” – Grupo B.....	184
Quadro 36 – Exemplos de palavras lexicais – Grupo B .....	185
Quadro 37 – Exemplos de palavras gramaticais – Grupo C.....	186
Quadro 38 – Exemplos de palavras lexicais – Grupo C .....	187
Quadro 39 – Exemplos de palavras gramaticais – Grupo D.....	188
Quadro 40 – Exemplos de palavras lexicais – Grupo D .....	188
Quadro 41 – Trechos de exemplo – palavras gramaticais em comum.....	190
Quadro 42 – Trechos de exemplo – palavras lexicais em comum .....	192
Quadro 43 – Exemplo de <i>hapax</i> específico de cada Grupo – vogal A .....	198
Quadro 44 – Exemplo de <i>hapax</i> específico de cada Grupo – consoante P .....	198
Quadro 45 – Exemplos – proposta 01 .....	201
Quadro 46 – Exemplos – proposta 03.....	202
Quadro 47 – Exemplos – proposta 05.....	202
Quadro 48 – Exemplos de uso – CorPop/ <i>corpus</i> de estudo.....	204

## LISTA DE TABELAS

Tabela 1 – Início da lista de itens lematizados do <i>corpus</i> de estudo.....	124
Tabela 2 – Redação por escola (primeiro estudo piloto) – texto, número de linhas e avaliação .....	126
Tabela 3 – Quantidade <i>tokens/types</i> – estudo piloto I.....	128
Tabela 4 – 20 palavras mais frequentes – textos dos alunos.....	129
Tabela 5 – 20 palavras mais frequentes – textos de contraste .....	130
Tabela 6 – Riqueza Lexical .....	131
Tabela 7 – Amostra de <i>hapax legomena</i> dos <i>corpora</i> analisados – vogal A.....	132
Tabela 8 – Amostra de <i>hapax legomena</i> dos <i>corpora</i> analisados – consoante P ...	133
Tabela 9 – Quantidade <i>tokens/types</i> – estudo piloto II.....	136
Tabela 10 – Relação de palavras lexicais e gramaticais – estudo piloto II.....	137
Tabela 11 – Riqueza lexical – estudo piloto II .....	138
Tabela 12 – <i>Hapax legomena</i> – estudo piloto II .....	139
Tabela 13 – <i>Tokens/types</i> – estudo piloto III.....	142
Tabela 14 – <i>Tokens/types</i> – proposta 3 .....	143
Tabela 15 – <i>Tokens/types</i> – proposta 5 .....	143
Tabela 16 – 50 primeiras palavras mais frequentes – estudo piloto III.....	144
Tabela 17 – Riqueza Lexical .....	146
Tabela 18 – <i>Tokens/types</i> – em ponderação – estudo piloto IV.....	152
Tabela 19 – Riqueza Lexical – estudo piloto IV.....	153
Tabela 20 – Relação <i>token/types</i> – experimento principal .....	160
Tabela 21 – Riqueza Lexical – experimento principal .....	162
Tabela 22 – Exemplos de riqueza lexical .....	179

## **APOIO DE FINANCIAMENTO CAPES**

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código 001.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>25</b>
1.1 ORIGEM DO TRABALHO, TEMA E OBJETIVOS .....	28
1.2 BREVE CARACTERIZAÇÃO DAS COMUNIDADES ESCOLARES ENVOLVIDAS .....	29
1.4 ORGANIZAÇÃO DO TRABALHO .....	32
<b>2 BASES TEÓRICAS E POSICIONAMENTO DO TRABALHO</b> .....	<b>34</b>
2.1 TRABALHOS RELACIONADOS .....	34
<b>2.1.1 Ensino de vocabulário em ensino de línguas</b> .....	<b>34</b>
<b>2.1.2 A produção textual como objeto de estudo: texto e vocabulário</b> .....	<b>37</b>
2.2 CONCEITOS-CHAVE E BASES TEÓRICAS .....	40
<b>2.2.1 A redação escolar como gênero textual e discursivo</b> .....	<b>40</b>
<b>2.2.2 Palavra</b> .....	<b>43</b>
<b>2.2.3 Léxico/Vocabulário</b> .....	<b>44</b>
<b>2.2.4 Palavras Gramaticais e Palavras Lexicais</b> .....	<b>50</b>
<b>2.2.5 Linguística de Corpus</b> .....	<b>53</b>
<b>2.2.6 Estatística Lexical nos estudos linguísticos</b> .....	<b>56</b>
<b>2.2.7 Frequência de palavras</b> .....	<b>60</b>
<b>2.2.8 Curvas de frequência, lematização e papel do hapax legomena</b> .....	<b>64</b>
<b>2.2.9 Riqueza Lexical</b> .....	<b>69</b>
<b>2.2.10 Keywords e chavicidade</b> .....	<b>72</b>
2.3 POSICIONAMENTO DO TRABALHO .....	78
<b>3 O ENSINO DE VOCABULÁRIO NAS AULAS DE LÍNGUA MATERNA NO ENSINO FUNDAMENTAL</b> .....	<b>91</b>
3.1 ENSINO DE VOCABULÁRIO .....	91
3.2 AULAS DE LÍNGUA MATERNA .....	95
3.3 ENSINO FUNDAMENTAL .....	99
<b>4 O VOCABULÁRIO NA ESCRITA E INTERSECÇÕES COM A LEITURA</b> .....	<b>104</b>
<b>5 EXPERIMENTO DESCRITIVO E ANALÍTICO</b> .....	<b>112</b>
5.1 MATERIAIS .....	113
5.2 ETAPAS DO EXPERIMENTO DESCRITIVO .....	120
5.3 PARTE I: ESTUDOS PILOTO .....	125
<b>5.3.1 Estudo Piloto I</b> .....	<b>126</b>

<b>5.3.2 Estudo Piloto II .....</b>	<b>135</b>
<b>5.3.3 Estudo Piloto III .....</b>	<b>140</b>
<b>5.3.4 Estudo Piloto IV.....</b>	<b>151</b>
5.4 PARTE II: EXPERIMENTO PRINCIPAL .....	159
5.5 KEYWORDS: AS PALAVRAS-CHAVE .....	167
5.6 SÍNTESE DOS RESULTADOS.....	171
<b>6 RETOMADA DAS QUESTÕES DE PESQUISA E HIPÓTESES .....</b>	<b>176</b>
6.1 QUESTÕES DE PESQUISA.....	176
6.2 RETOMADA DAS HIPÓTESES .....	189
6.3 SÍNTESE .....	207
<b>7 SUGESTÕES DE ATIVIDADES DE ENSINO DE VOCABULÁRIO .....</b>	<b>210</b>
7.1 PROPOSTA DE UNIDADE DE ENSINO DE VOCABULÁRIO.....	211
<b>7.1.1 Unidade de Ensino de Vocabulário.....</b>	<b>212</b>
7.2 PROPOSTAS DE EXERCÍCIOS (INDEPENDENTES) COM VOCABULÁRIO... 218	
<b>7.2.1 Reescritura e contexto.....</b>	<b>220</b>
<b>7.2.2 Relação entre vocábulo e sentido.....</b>	<b>221</b>
<b>7.2.3 Criação de contextos para determinados sentidos.....</b>	<b>222</b>
<b>7.2.4 Exploração de sentido de acordo com o contexto.....</b>	<b>222</b>
<b>7.2.5 Conhecido X Desconhecido .....</b>	<b>223</b>
<b>7.2.6 Variação de formas .....</b>	<b>224</b>
<b>7.2.7 Palavras de diferentes origens I.....</b>	<b>225</b>
<b>7.2.8 Palavras de diferentes origens II.....</b>	<b>226</b>
<b>7.2.9 Acessibilidade textual.....</b>	<b>227</b>
<b>7.2.10 Estatística Lexical .....</b>	<b>228</b>
7.3 COMENTÁRIOS SOBRE AS PROPOSTAS .....	230
<b>8 CONSIDERAÇÕES FINAIS: LIMITES E PERSPECTIVAS.....</b>	<b>233</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>238</b>
<b>ANEXO I – UMA REDAÇÃO COMPLETA DE CADA ESCOLA .....</b>	<b>247</b>
<b>ANEXO II – TEMAS E PROPOSTAS DAS PRODUÇÕES.....</b>	<b>249</b>
<b>ANEXO III – CORPORA DE CONTRASTE: TEXTOS DE APOIO .....</b>	<b>250</b>
<b>ANEXO IV – CORPORA DE CONTRASTE: TEXTOS DOS LIVROS DIDÁTICOS</b>	<b>253</b>
<b>ANEXO V – FILMES USADOS COMO “APOIO” (RESUMO, IMAGEM) .....</b>	<b>256</b>
<b>ANEXO VI – MATERIAL DE APOIO À UNIDADE DE ENSINO: FIGURAS E TEXTOS.....</b>	<b>257</b>

<b>ANEXO VII – MATERIAL DE APOIO À UNIDADE DE ENSINO: QUESTÕES DE INTERPRETAÇÃO.....</b>	<b>261</b>
<b>ANEXO VIII – MATERIAL DE APOIO À UNIDADE DE ENSINO: PROPOSTA DE PRODUÇÃO TEXTUAL.....</b>	<b>262</b>
<b>ANEXO IX – MATERIAL DE APOIO À UNIDADE DE ENSINO: MODELO DE RELATÓRIO .....</b>	<b>263</b>
<b>ANEXO X – MATERIAL DE APOIO À UNIDADE DE ENSINO: RIQUEZA LEXICAL .....</b>	<b>264</b>
<b>ANEXO XI – MATERIAL DE APOIO À UNIDADE DE ENSINO: TEXTOS COM DIFERENTES RIQUEZAS .....</b>	<b>265</b>
<b>ANEXO XII – MATERIAL DE APOIO À UNIDADE DE ENSINO: MODELO DE GRÁFICO.....</b>	<b>267</b>
<b>ANEXO XIII – MATERIAL DE APOIO À UNIDADE DE ENSINO: AUTOAVALIAÇÃO .....</b>	<b>268</b>
<b>ANEXO XIV – <i>WORDLIST</i> DO CORPUS DE ESTUDO .....</b>	<b>270</b>



## 1 INTRODUÇÃO

Esta pesquisa de mestrado, iniciada em agosto de 2018, situa-se no âmbito dos estudos da Lexicologia. Nossos objetivos são descrever e analisar o vocabulário escrito empregado por estudantes do Ensino Fundamental da rede pública de educação da cidade de Porto Alegre – RS. O vocabulário escrito, o qual buscamos apreender por meio do estudo de uma amostra, é aquele empregado em um gênero textual e discursivo típico da sala de aula: a redação escolar.

A ideia base desta pesquisa é contribuir para que se possa traçar um perfil de vocabulário reconhecido e manifestado por esses estudantes desta etapa do ensino. Também se espera contribuir para minimizar uma lacuna de conhecimento, visto que esse tema tem sido pouco explorado em pesquisas de pós-graduação.

Para tanto, reunimos, em um *corpus*, 154 produções textuais de alunos de 6º ano e de 8º ano de duas escolas públicas: uma do Município de Porto Alegre e outra do Governo do Estado do Rio Grande do Sul. As redações sob exame, manuscritas e de temáticas diversas, fazem parte de um acervo pessoal, desidentificado, colecionado pela professora de Língua Portuguesa dessas duas escolas, ao longo dos seus oito anos de docência.

Os textos do *corpus* foram todos produzidos em sala de aula, num período de cerca de 90 minutos, equivalente a dois períodos, de acordo com as orientações e propostas de produção. Alguns tiveram como base textos de apoio, outros, filmes assistidos também na escola, durante as aulas de Língua Portuguesa.

Com a finalidade de contrastar o vocabulário empregado nessas redações, inicialmente utilizamos textos dos livros didáticos de Língua Portuguesa que foram trabalhados com os alunos nessas duas escolas em período anterior à escrita dos textos. São eles: *Tecendo Linguagens* (OLIVEIRA *et al.*, 2015) e *Português – Linguagens* (CEREJA, MAGALHÃES, 2014). Esse contraste foi pensado com o objetivo de verificar qual a influência dos textos dos livros sobre o vocabulário escrito utilizado pelos alunos nas suas produções.

Além disso, recorreremos também aos textos de apoio ou textos motivadores apresentados aos alunos no momento em que escreveram as suas redações. Neste caso, novamente, buscamos verificar o quanto haveria de influência ou de reutilização de palavras. Assim, quando existentes, mapeamos o léxico desses materiais que motivaram as produções escritas. A seguir, observamos trecho de

uma das redações do *corpus* de estudo e de um dos textos de apoio que compõem esse *corpus* de contraste:

Quadro 1 – Trecho de redação/ trecho de texto de apoio

Trecho de redação ( <i>corpus</i> de estudo)	Trecho do texto de apoio 'A vida como filme' – Moacyr Scliar ( <i>corpus</i> de contraste)
Se eu pude-se fazer um filme da minha vida ia ser filme de comédia, tragédia e drama, ação... Pois minha vida é agitada, iria ser muito engraçado porque eu iria contar todos os meus “podres” da minha vida.	Se pudéssemos fazer de nossa vida um filme seria, não tenhamos dúvida, um grande filme. Uma super-produção, dessas que custam US\$100 milhões e batem recordes de bilheteria.

Fonte: Acervo pessoal.

Ainda como contraste, mas, neste caso, para verificação do vocabulário dos estudantes frente a um quadro geral do português escrito atual, recorreremos a *corpora* de referência. Para esse papel referencial, selecionamos o CorPop (PASQUALINI, 2018), um *corpus* de referência do português popular escrito culto<sup>1</sup> do Brasil. Ele é composto por materiais de literatura considerados mais acessíveis para leitores adultos, de baixa escolaridade, com letramento básico e pertencentes às classes C e D; e por textos de jornais populares consumidos por esse público. A ideia do CorPop era reunir um *corpus* de textos “simples” do português popular brasileiro, em que “*Simples*, aqui, significa, grosso modo, um texto que seja compatível com o nível de letramento médio da maioria dos brasileiros adultos e, a princípio, não seja de áreas especializadas” (PASQUALINI, 2018, p. 14).

Além dele, o Banco de Português, o Lácio Web e o Corpus brasileiro também foram utilizados como *corpora* de referência do Português, ao analisamos as palavras-chave do *corpus* de estudo.

Ao iniciar o trabalho na Rede Municipal, eu já trazia a experiência da Rede Estadual. Em função disso, logo percebi diferenças entre os estudantes e suas competências leitoras. Os alunos do Município, aparentemente, apresentam mais dificuldades de aprendizagem e, sobretudo, de leitura, de interpretação e de compreensão de textos e de questões, além de falhas de ortografia, de formação de palavras e de frases, de concordância, entre outros. Isso, conforme o grupo de

<sup>1</sup> As noções *norma culta* e *norma padrão* correspondem a conceituações complexas. O *corpus* em questão reúne desde textos jornalísticos feitos por pessoas em situação de rua até peças literárias adaptadas por linguistas para pessoas com pouca escolaridade. Sem entrar na discussão, optamos, aqui, pela denominação *norma culta*.

professores que lecionava para os dois diferentes grupos, tendia a afetar seus desempenhos não só nas aulas de Língua Portuguesa, mas também de todas as outras disciplinas. A partir dessa impressão subjetiva, busquei construir um entendimento objetivo e cientificamente amparado para a situação vivenciada. Assim nasceu esta pesquisa, a qual, por uma contingência natural, limitar-se-á a tratar, em contrastes, do comportamento do vocabulário em meio à escrita dos dois grupos de alunos.

Então, com auxílio da estatística lexical (BIDERMAN, 1978; 1998) e da Linguística de Corpus (BERBER SARDINHA, 2004), o objetivo principal deste trabalho é descrever e analisar esse vocabulário escrito. Como um objetivo secundário, buscamos identificar se as diferenças subjetivas percebidas pelos professores das duas redes – municipal e estadual – existiriam, de fato, do ponto de vista do léxico.

Além disso, outro objetivo associado ao trabalho é subsidiar uma proposta de unidade de ensino especialmente dedicada à temática do vocabulário, baseada nos padrões descritos e analisados neste estudo. Acreditamos que a pesquisa científica não só pode e deve contribuir para que se possa refletir sobre a atividade do profissional docente, acolhendo suas experiências na academia, como também pode e deve estar a serviço de propostas pedagógicas que sejam executáveis em sala de aula, objetivando a aprendizagem do estudante.

De acordo com Silva (2016, p. 122), “é nesse caso que a LC se torna uma importante aliada, pois é por meio dela que podemos estabelecer padrões advindos da língua alvo e criar materiais que correspondam às necessidades dos alunos”. Nessa proposta, trazemos sugestões de atividades que podem ser aplicadas pelos professores em sala de aula. Acreditamos que tais atividades tendem a contribuir com o crescimento e com as aprendizagens do aluno, sobretudo no que diz respeito ao desenvolvimento de sua proficiência leitora, por meio do alargamento do seu vocabulário.

A partir da digitação dos textos manuscritos, com a devida preservação das grafias originais e de eventuais erros, organizamos uma base de dados. A seguir é possível visualizar um trecho original de uma das redações e sua respectiva digitação em arquivo de texto:

Figura 1 – Trecho de texto original manuscrito I

s muito legais, Realmente teve coisas que o Poderiam não ter acontecido, como o dia 7 que pisei no formigueiro e tive que correr para a praia, Realmente foi hilário.

Fonte: Acervo pessoal.

Quadro 2 – Trecho de texto original digitado I

<p>muito legais. Realmente teve coisas que poderiam não ter acontecido, como o dia que pisei no formigueiro e tive que correr para a praia, realmente foi hilário.</p>
--

Fonte: Acervo pessoal.

Com essas redações digitadas, agrupamos todos os arquivos em bloco único, conforme a escola/rede, conforme a proposta de produção e conforme a etapa escolar dos alunos que as escreveram. Para explorar nossa base de dados, utilizamos uma ferramenta computacional que apoiou nossa descrição estatística do léxico. A ferramenta, chamada *AntConc* (ANTHONY, 2019), também usada para a verificação de vocabulário nos materiais de contraste com as redações, é um *software* de acesso livre, que auxilia na análise estatística descritiva dos textos e que realiza a quantificação das palavras neles contidas.

Com esses dados, foi possível contrastar os resultados das duas escolas entre si. Além disso, os contrastamos com os materiais de apoio (que serviram como contraste), e com os *corpora* de referência.

## 1.1 ORIGEM DO TRABALHO, TEMA E OBJETIVOS

Conforme já citado, este trabalho de investigação teve origem nas inquietações de docentes de duas escolas públicas do Rio Grande do Sul: uma municipal e outra estadual. As dificuldades, em diferentes níveis – de leitura, de interpretação e de compreensão dos alunos –, pareciam ultrapassar as aulas de Língua Portuguesa. Professores de outras áreas, que lecionavam nessas duas realidades, também relatavam a situação. O que lhes parecia, na época, era que os alunos da Rede Municipal, ainda que contassem com maiores recursos e melhor estrutura, mostravam desempenhos inferiores aos dos alunos da Rede Estadual.

Acreditamos que tal percepção de diferença entre os desempenhos de compreensão e de leitura entre os alunos dos dois grupos poderia ser explicada –

justificada ou negada – por meio da observação atenta dos diferentes repertórios vocabulares presentes nas suas produções escritas. Assim, a aposta da nossa investigação é de que as produções escritas dos alunos, a partir do vocabulário que utilizam, são capazes de subsidiar uma melhor compreensão sobre eventuais diferenças de desempenhos e de aproveitamento dos dois grupos.

Nesse sentido, o tema específico deste trabalho corresponde aos padrões de vocabulário escrito desses alunos. A partir da descrição desse vocabulário, buscamos entender um pouco melhor o desenvolvimento do estudante de Ensino Fundamental.

Por fim, pensamos que essa descrição de vocabulário escrito tal como empregado pelos alunos poderá subsidiar uma proposta de material didático com sugestões de atividades capazes de influenciar positivamente as suas capacidades de interpretação e de compreensão. Por meio da expansão do seu vocabulário ativo, acreditamos que o aluno tenderá a uma competência leitora também ampliada. A ideia é contribuir com o alargamento do repertório lexical que os alunos já têm, bem como “para a conscientização do seu uso adequado e para a sua correção quanto a aspectos morfológicos, sintáticos e [...] semânticos” (BARBEIRO, [entre 2003 e 2007], p. 01).

## 1.2 BREVE CARACTERIZAÇÃO DAS COMUNIDADES ESCOLARES ENVOLVIDAS

As escolas cujas vivências foram aproveitadas como pano de fundo para este trabalho situam-se na Zona Norte de Porto Alegre, capital do Rio Grande do Sul, ambas no mesmo bairro, cerca de 3 km de distância uma da outra. A distância entre as duas escolas também foi um fator que interessou para a delimitação da pesquisa, já que se localizam muito próximas, de modo que o fator localização, a princípio, não interferiria nos resultados da pesquisa. O fato de uma escola ser estadual e, a outra, municipal, conforme já reiterado, é algo a ser inserido no cenário de pesquisa, considerando-se as duas realidades de ensino público.

A primeira escola que fornece dados de observação contextual é uma escola de Ensino Fundamental administrada pelo Governo do Estado do Rio Grande do Sul, fundada em 1943. Entre os anos de 2017 e 2018, nos quais foram produzidos as redações sob exame, a escola atendia cerca de 430 alunos. Lecionei nessa escola, nas turmas de 6º a 9º ano do Ensino Fundamental, desde o ano de 2013 até

o ano de 2018. A escola é organizada por séries/anos, tem no máximo duas turmas por série/ano, conta com 19 professores (dos quais a maioria tem Pós-Graduação ou Especialização) e divide uma professora especializada no atendimento de alunos com necessidades especiais com outras escolas do mesmo bairro.

A segunda é a escola de Ensino Fundamental administrada pela Prefeitura de Porto Alegre, fundada em 1963. Entre os anos 2018 e 2019, nos quais foram produzidos as redações sob exame, a escola atendia cerca de 924 alunos. Nessa escola, tenho lecionado, de 2016 a 2021, nas turmas de primeiro ano e de segundo ano do terceiro ciclo, respectivamente, C10 e C20. A escola é organizada por ciclos de formação, possui pelo menos três turmas de cada ano/ciclo e conta com um quadro de 64 professores (dos quais a maioria tem Pós-Graduação ou Especialização, muitos têm Mestrado e alguns têm Doutorado em sua área de formação). Na época da produção das redações sob exame, apresentava projetos educacionais de cooperação como FECI (Fundação de Educação e Cultura do Sport Club Internacional), LIAU (Laboratório de Inteligência do Ambiente Urbano) ou LENUM (Letramento e Numeramento), atualmente desativados. A escola segue participando de programas como Adote um Escritor e dispõe de Laboratório de Aprendizagem e de Sala de Integração e Recursos (SIR), com duas professoras especializadas no atendimento de alunos com necessidades especiais.

Quanto ao perfil socioeconômico dos alunos das duas escolas, cabe destacar, em linhas gerais, que a maioria residia próximo às escolas, podendo deslocarem-se a pé para as aulas. Assim, temos duas escolas de bairro de periferia atendendo estudantes residentes nesse bairro. As famílias dos estudantes eram bastante presentes na comunidade escolar a despeito de eventuais dificuldades financeiras.

Os alunos pesquisados tinham entre 11 e 15 anos, sendo a média de idade dos dois grupos 13 anos. Alguns desses alunos eram repetentes, porém, a maioria (cerca de 80%) estava cursando a série recomendada para sua faixa etária. A maioria dos estudantes, nas duas escolas, no período do estudo, mostrava-se bastante envolvida com as atividades propostas, incluindo a produção de texto.

### 1.3 QUESTÕES DE PESQUISA E HIPÓTESES

Conforme já mencionado, as duas escolas são públicas e localizadas na mesma região da capital. Ainda assim, logo que iniciamos o trabalho na Rede Municipal, percebemos que havia grande diferença entre os cenários, as práticas e seus sujeitos. Os alunos da Rede Municipal, aparentemente, apresentavam maiores dificuldades (sobretudo de leitura, de interpretação e de compreensão) que os alunos da Rede Estadual.

Em função disso e de lecionar a mesma disciplina (Língua Portuguesa), paralelamente, nas duas redes, para os mesmos anos/ciclos (que se correspondem, nos dois modelos de educação), surgiu meu interesse em observar e pesquisar essa situação de modo científico. Então, este trabalho pretende responder às seguintes perguntas:

- a) Os alunos dessas duas redes de ensino público apresentam diferentes perfis de vocabulário escrito?
- b) Quais elementos de natureza estatística permitem caracterizar o vocabulário escrito dos dois públicos?
- c) Quais elementos de natureza estatística permitem distinguir o vocabulário de um grupo em relação ao outro?
- d) Como são os perfis de vocabulário escrito dos diferentes grupos de redatores envolvidos na amostra?

A partir dessas questões, as hipóteses deste trabalho são as seguintes:

- a) Os alunos dos dois principais grupos pesquisados apresentam diferenças quanto ao perfil do vocabulário escrito no que se refere a(s):
  - I) repertório de formas gramaticais mais frequentes;
  - II) repertório de formas lexicais mais frequentes;
  - III) variedade do vocabulário;
  - IV) especificidade(s) do vocabulário.
- b) Os alunos apresentam influência do vocabulário do(s) texto(s) de apoio nas suas redações.
- c) O vocabulário escrito dos alunos é próximo de um padrão de português popular escrito, como o representado pelo CorPop.

## 1.4 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado da seguinte maneira: no primeiro capítulo, apresentamos os antecedentes da pesquisa, contextualizando o cenário envolvido, os alunos e as escolas. Nesse sentido, também descrevemos a origem do trabalho e os motivos que nos levaram a explorar a temática do vocabulário escrito dos alunos da rede pública de educação, contrastando as duas redes nas quais leciono: municipal e estadual. Ainda nesse capítulo introdutório, apresentamos nossos objetivos, nossas hipóteses e as questões de pesquisa que movem o nosso estudo.

No segundo capítulo, são trazidos nossos referenciais teóricos. Primeiramente, discorreremos sobre estudos já realizados sobre redações de vestibular, Exame Nacional do Ensino Médio (Enem) e exames de proficiência, demonstrando como essas redações são mais estudadas e pesquisadas que as produções de alunos de Ensino Fundamental. Ainda existem lacunas nos estudos com esses textos, desenvolvidos em menor quantidade, conforme nossas buscas e pesquisas aos trabalhos já realizados. Após, apresentamos as bases teóricas que guiam nossa pesquisa, como a Linguística de Corpus e a Estatística Lexical, bem como as noções de palavra, léxico e vocabulário, colocando, ao final, o nosso posicionamento teórico adotado neste trabalho.

Nos terceiro e quarto capítulos, são focados no ensino de vocabulário nas aulas de língua materna da escola pública. No terceiro capítulo, traçamos um panorama sobre ensino, ensino de vocabulário, aulas de língua materna e Ensino Fundamental. Já no quarto capítulo, mostramos a intersecção desse panorama com a leitura e a escrita em sala de aula, que resultaram nas produções escritas dos alunos analisadas neste trabalho.

No quinto capítulo, descrevemos nossos materiais e nossos métodos de pesquisa. O *corpus* de estudo e os *corpora* (tanto de contraste, quanto de referência) são discriminados, bem como os procedimentos realizados. Em seguida, relatamos nossos estudos piloto, com partes específicas do nosso *corpus* total de estudo, que foram aumentando a cada piloto. Essas explorações iniciais contribuíram para a delimitação dos nossos materiais e métodos, que foram sendo ajustados ao longo do percurso e que resultaram no experimento principal, também apresentado e detalhado nesse capítulo. Ao final do capítulo, manifestamos nossas considerações acerca dos experimentos realizados.



O sexto capítulo traz a retomada das questões de pesquisa e das hipóteses do trabalho, conforme o que foi apontado em nossos experimentos. Esses resultados possibilitaram a produção de uma unidade de ensino de vocabulário, bem como de atividades individuais, idealizadas para aplicação em sala de aula de escola pública, no Ensino Fundamental. Essa unidade e essas atividades são apresentadas no sétimo capítulo. A intenção é de que sejam compartilhadas com os colegas, professores de língua materna. O objetivo de criar tais propostas foi o de auxiliar na expansão do vocabulário escrito dos alunos, pois acreditamos que quanto maior sua variedade de vocabulário, maior também será sua proficiência em leitura, compreensão e interpretação, questões primordiais que deram origem a este trabalho.

Por fim, o oitavo e último capítulo tece algumas considerações finais sobre nosso estudo. Discorreremos sobre os limites enfrentados nesta pesquisa, sobre quais deles conseguimos ultrapassar e sobre quais permaneceram, tornando-se perspectiva de estudo futuro.

## **2 BASES TEÓRICAS E POSICIONAMENTO DO TRABALHO**

Neste capítulo, expomos os trabalhos relacionados ao nosso e trazemos os referenciais teóricos que guiam esta investigação. Primeiramente, apresentamos um panorama geral de estudos já realizados nas áreas com as quais o nosso trabalho se relaciona, como o ensino de vocabulário e as produções de textos por parte dos alunos (redações de vestibular, Enem e exames de proficiência, por exemplo). Em seguida, discutimos os principais conceitos que nosso estudo mobiliza, como as noções de palavra, léxico e vocabulário, entre outros. Na sequência, tratamos de áreas de estudos como a Linguística de Corpus e a Estatística Descritiva Lexical. Ao final do capítulo, trazemos o posicionamento teórico adotado neste trabalho.

### **2.1 TRABALHOS RELACIONADOS**

Apresentamos aqui trabalhos que julgamos mais relevantes, observando nossos objetivos e nosso objeto de estudo, sem pretensão de exaustividade. Os primeiros discutem o ensino de vocabulário no ensino de línguas. Depois, revisamos pesquisas que têm a produção textual como objeto de estudo, sejam redações escolares (de Ensino Fundamental ou Médio), de vestibular, do Enem ou de exames de proficiência.

#### **2.1.1 Ensino de vocabulário em ensino de línguas**

O ensino de vocabulário desempenha papel importante no ensino de línguas, pois, assim como bem coloca Serra (2016, p. 02): “Não se pode esquecer que, muitas vezes, os problemas de leitura e escrita que o aluno enfrenta em seu cotidiano passam pela falta do conhecimento de um vocabulário amplo, além daquele que ele traz de sua comunidade de fala”. Em função de observar tais dificuldades em nossa prática docente, tal temática é abordada em nosso trabalho.

Conforme Colombo e Cárnio (2018, p. 02), “embora existam estudos brasileiros sobre vocabulário e outras habilidades linguísticas, poucos avaliaram a relação entre vocabulário e compreensão de leitura textual”. Essa é, então, a nossa ideia, pois acreditamos que as dificuldades de leitura e de compreensão dos alunos tendem a estar espelhadas no repertório ou no perfil do seu vocabulário escrito.

Ao pesquisarmos no Google Acadêmico e em bases de dados de bibliotecas digitais acadêmicas, como a SciELO,<sup>2</sup> obtemos como resultado muitos trabalhos sobre o ensino de vocabulário voltado ao ensino de línguas estrangeiras. Ao que nos parece, o ensino de língua estrangeira está conectado ao ensino de vocabulário, como podemos ver nestes dois recentes trabalhos que tratam do assunto.

Pires e Tumolo (2020) analisaram o desenvolvimento de vocabulário em três aplicativos para aprendizagem de língua inglesa. Seus resultados mostraram que os aplicativos instruem as formas oral e escrita das palavras, porém utilizam imagens e vídeos que nem sempre são relevantes. Em função disso, concluíram que os aplicativos podem auxiliar no desenvolvimento de vocabulário da língua, ainda que tenham algumas limitações.

Já Carecho, Soares, Fernandes (2020) descreveram uma experiência de aprendizagem de língua alemã, por meio de recursos desenvolvidos para aprendizagem autônoma, utilizados por estudantes de alemão da Faculdade de Letras da Universidade de Coimbra. A avaliação realizada foi positiva e os alunos “[...] consideraram os recursos disponibilizados úteis para sua aprendizagem, já que esta era a única compensação para a participação no projeto” (CARECHO, SOARES, FERNADES, 2020, p. 272).

No entanto, no contexto do ensino de nossa língua materna, a Língua Portuguesa, o ensino de vocabulário parece ser pouco pesquisado cientificamente, conforme já apontado por Ferreira e Vieira (2013, p. 19): “[...] ficou evidenciado que o estudo do vocabulário/léxico é notadamente voltado para o ensino de língua estrangeira, o que tornam escassas as pesquisas sobre o ensino de vocabulário/léxico nas aulas de língua materna”.

Ferreira e Vieira (2013) realizaram pesquisa bibliográfica sobre a importância do estudo de vocabulário/léxico nas aulas de língua materna. Elas trataram de bases históricas, abordagem pedagógica e considerações teórico-metodológicas sobre o assunto.

Antes dessa percepção dos autores anteriores, encontramos o estudo de Dargel (2011), que salientou que seu trabalho foi motivado justamente pela

---

<sup>2</sup> A *Scientific Electronic Library Online* – SciELO – é uma biblioteca eletrônica que abrange uma coleção selecionada de periódicos científicos brasileiros. O Projeto tem por objetivo o desenvolvimento de uma metodologia comum para a preparação, armazenamento, disseminação e avaliação da produção científica em formato eletrônico.

consciência da importância que a ampliação lexical tem na vida e na caminhada de um ser humano. Nesse sentido, sua tese apresentou subsídios metodológicos, teóricos e práticos - como orientações teóricas sobre o ensino do vocabulário, sugestões sobre como ensinar o aluno a conhecer e a gostar de usar o dicionário, exercícios diversos a respeito do vocabulário - que auxiliam o professor de Língua Portuguesa do Ensino Fundamental II.

Já Pais e Sardinha (2011) relataram, em seu trabalho, a relação direta existente entre ensino explícito de vocabulário, prática de leitura e sucesso escolar. Os autores apresentaram resultados de sucesso de pesquisas realizadas no exterior que interligaram essas três vertentes, desde que o ensino de vocabulário não fosse descontextualizando e não visasse apenas à memorização.

No mesmo caminho, Serra (2016) explorou a temática do ensino de vocabulário nas escolas do Brasil, avaliando sua importância para produção e coesão do texto. O autor considerou que o ensino de vocabulário contextualizado resulta na melhora da habilidade de produção e leitura de textos.

Hartmann (2020), de maneira inovadora, trabalhou com a simplificação de textos informativos para que fossem mais bem compreendidos por alunos de Ensino Fundamental. O objetivo de sua tese foi o de alavancar a área de pesquisa em Adaptação Lexical na Língua Portuguesa, atuando na Simplificação e Elaboração Lexical, a fim de adaptar textos autênticos para um determinado nível escolar. Essa adaptação permitiria que um texto de maior complexidade fosse lido pelos alunos, que, assim, conseguiriam compreender seu conteúdo. Isso tudo foi realizado de maneira informatizada, por cientistas da computação. Além disso, também foi realizada testagem com os alunos do Projeto Pequeno Cidadão.<sup>3</sup>

Em comum, os trabalhos aqui citados concordam com a importância do ensino de vocabulário para as aulas de língua materna. Colombo e Cárnio (2018, p. 02) inclusive defendem que “[...] o vocabulário adquirido durante a leitura e os conhecimentos prévios do leitor contribuem para a compreensão de leitura”. Além disso, assim como nosso estudo, os autores referidos salientam a ligação que um vocabulário extenso exibe com as competências de leitura, interpretação,

---

<sup>3</sup> O Projeto Pequeno Cidadão, da Universidade de São Paulo, campus São Carlos, desde 1996 trabalha para promover a cidadania com formação educacional e profissional para crianças. Proporciona aulas de reforço (Português e Matemática) para crianças de famílias de baixa renda, além de diversas oficinas. Disponível em: <<http://pequenocidadao.sc.usp.br/>>. Acesso em: 27 abr. 2020.

compreensão e escrita de um indivíduo, sobretudo de crianças, jovens e adolescentes em formação.

### **2.1.2 A produção textual como objeto de estudo: texto e vocabulário**

Sabemos que o ensino de vocabulário tem impacto nas produções de textos realizadas pelos alunos em sala de aula. Quanto maior o vocabulário de um indivíduo, melhor ele conseguirá se expressar, argumentar e construir seu texto de maneira coerente e coesa. Em função disso, nossa pesquisa relaciona essas duas temáticas. Acreditamos na importância de um indivíduo ter um vocabulário variado, pois sabemos que esse fator auxilia na coesão e na produção de um texto. Então, além de nos preocuparmos com o ensino de vocabulário, abordamos também outra temática pouco explorada: as produções de alunos de Ensino Fundamental.

Ao pesquisarmos na base de dados Sabi+,<sup>4</sup> percebemos que boa parte dos estudos realizados com textos do tipo “redação escolar” ocupam-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). Ao que parece, pouca atenção se dá às produções escolares de alunos de Ensino Fundamental.

Endruweit (2000), por exemplo, pesquisou sobre a evolução das temáticas da redação do vestibular da UFRGS. Em sua dissertação, a autora apresenta a pressão que esse assunto sempre exerceu sobre os alunos do Ensino Médio, bem como a análise do que leva uma redação a ser considerada nota 10 (conforme o próprio título da sua dissertação).

Ainda sobre redações do vestibular da UFRGS, Evers (2018) descreveu e analisou 341 redações do concurso de 2014, cujo tema da proposta foi “O meu clássico”. Sob viés estatístico, ela também investigou quais foram os quesitos contidos em cada redação, separando-os de acordo com faixas de notas. A autora considerou itens como a quantidade total de palavras, de sentenças e de parágrafos; os desvios de ortografia; a pontuação e o índice de complexidade textual, medido

---

<sup>4</sup> Catálogo on-line – SABi+ UFRGS. O Sistema de Automação de Bibliotecas foi implantado em 1989 e adota o software Aleph 500 para gerenciar as atividades e serviços oferecidos pelas 33 bibliotecas da UFRGS a sua comunidade usuária. Contém o Lume-UFRGS, o Portal de Periódicos da Capes, entre outras assinaturas da UFRGS.

pelo Índice Flesch.<sup>5</sup> A avaliação dessas características presentes nos textos possibilitou a estruturação de um padrão de redação de vestibular a partir do qual os alunos do cursinho pré-vestibular popular pesquisado seriam capazes de atingir a aprovação nos cursos requeridos.

No Estado de São Paulo, as redações de vestibular também foram pesquisadas. Nascimento e Isquierdo (2003) analisaram 450 redações de vestibular, dos anos 1999 e 2000, de duas universidades do Estado, uma pública e outra privada. Os dados obtidos foram contrastados com os dados disponíveis do *Dicionário de frequências do léxico do português*, de Biderman (1998a), e apontaram para um possível núcleo de vocabulário, que seria utilizado em qualquer gênero textual.

Luiz (2018) investigou as redações “estilo Enem” produzidas por alunos de Ensino Médio de uma escola particular de Uberlândia (MG). A autora percebeu que os estudantes não conseguiam compreender a temática a ser desenvolvida nesse gênero textual, nem associar repertório sociocultural ao tema proposto ou estruturar orações, frases e parágrafos de forma a apresentar um posicionamento crítico. Em função disso, construiu um planejamento didático focado na composição da prova de redação do Enem. Sua curiosidade era analisar o desenvolvimento das capacidades escrita, crítica e argumentativa dos alunos não só para esses exames de redação, mas também para vida.

É possível encontrar outros trabalhos centrados nos mesmos tipos de produções textuais: redação de vestibular e/ou Enem. Além desses, também encontramos estudos com material escrito por estudantes de Português como língua estrangeira.

Evers (2013) pesquisou os textos produzidos no exame que confere o Certificado de Proficiência em Língua Portuguesa para Estrangeiros (Celpe-Bras) de 2006-1. O Celpe-Bras é um exame de proficiência em Língua Portuguesa, conferido aos estrangeiros que têm a necessidade de se integrar à vida no Brasil ou de usar o português no exterior. Com enfoque computacional e pautado na léxico-estatística, a autora investigou a hipótese de classificar, de modo automático, textos submetidos

---

<sup>5</sup> Índice estatístico que indica o grau de complexidade de um texto. No Brasil, a fórmula de Flesch foi adaptada por pesquisadores do Núcleo Interinstitucional de Linguística Computacional da USP e da UFSCar para o português brasileiro (MARTINS *et al.*, 1996), bem como a ferramenta Coh-Metrix (SCARTON; ALUÍSIO, 2010).

ao exame conforme níveis de proficiência pré-estabelecidos. Para tanto, analisou itens como Índice Flesch, riqueza lexical dos textos, entre outros. Um dos seus resultados apontou que o número de níveis era grande demais para que a máquina fosse capaz de distinguir sozinha os textos a partir dos itens identificados, mostrando que ainda é preciso trilhar um longo caminho para a automatização dessa tarefa.

Com relação a produções textuais de alunos de 5º ano, considerado Ensino Fundamental I, Fortes (2016) pesquisou o processo de reescrita de textos pelos próprios alunos escritores. Seu estudo observou as mudanças de grafia, pontuação e usos de elementos coesivos e sequenciais de uma versão do texto para a outra. Sua ideia foi verificar o quanto essa revisão que o aluno autor faz do seu texto auxilia no avanço de seu processo de escrita e de produção textual.

Os trabalhos até aqui citados foram localizados em pesquisa bibliográfica no Scielo e no SABI+, com vistas a colocarmos, nesse tópico de pesquisa, o estado da arte. Entretanto, em busca no sistema Lume-UFRGS,<sup>6</sup> não encontramos trabalhos relacionados a textos de alunos de Ensino Fundamental II (do 6º ao 9º ano). Isso confirma a tendência de serem pouco explorados em pesquisas acadêmicas. Nesse sentido, nossa ideia é suprir tal lacuna de estudos, afinal, é com a “bagagem” de saberes e competências do Ensino Fundamental que os estudantes enfrentam os desafios do Ensino Médio e, em seguida, dos exames de redação, como vestibular e Enem. Então, pretendemos contribuir, sob a perspectiva da léxico-estatística descritiva, para que a produção textual da segunda etapa do Ensino Fundamental também seja explorada.

Após essa breve caracterização de trabalhos já realizados e relacionados ao nosso, a próxima seção deste capítulo apresenta os conceitos-chave e os referenciais teóricos envolvidos nesta pesquisa. Em seguida, evidenciamos os posicionamentos específicos adotados em nosso estudo.

---

<sup>6</sup> Repositório digital da Universidade Federal do Rio Grande do Sul. Portal de acesso às coleções digitais produzidas no âmbito da Universidade e de outros documentos de interesse da instituição. Contém teses, dissertações, trabalhos de conclusão de curso de mestrado profissional ou especialização, além de trabalhos de conclusão de curso de graduação.

## 2.2 CONCEITOS-CHAVE E BASES TEÓRICAS

Nesta seção, abordamos os principais conceitos e teorias nos quais nos baseamos para a construção do nosso estudo. Realizamos uma revisão da bibliografia da área, apresentando ideias de diferentes autores e fazendo ponderações acerca das correlações com o nosso estudo.

Tais conceituações foram importantes para apoiar os estudos piloto e o experimento principal de nossa pesquisa. Muitos dos conceitos aqui revisados são retomados no quinto capítulo deste trabalho, no qual tratamos de materiais e métodos.

### 2.2.1 A redação escolar como gênero textual e discursivo

De acordo com Antunes (2003), “toda atuação verbal se dá através de textos, independente de sua função e de sua extensão” (ANTUNES, 2003, p. 92). Ou seja, a autora defende que, aos nos comunicarmos, seja de maneira oral, seja de maneira escrita, o fazemos por meio de textos. Por isso, sustenta que um texto tem diferentes formas, funções e extensões. Segundo a autora:

Fazer e entender textos não é assim uma atividade eventual, alguma coisa que as pessoas fazem uma vez ou outra, em circunstâncias muito especiais (em dias de prova, por exemplo). É coisa que se faz todo dia, *sempre* que se fala ou *sempre* que se escreve. (ANTUNES, 2003, p. 92).

De certa forma, anterior, porém consoante às ideias de Antunes, Bakhtin (2003) acreditava que só falamos, escrevemos e nos comunicamos por meio de gêneros do discurso. Para o autor, cada um de nós tem um repertório infinito de gêneros e o mobilizamos sem sequer percebermos.

Então, a partir dessas ideias, Antunes (2003) salienta o lugar e a importância da produção de textos na escola. Como é papel da escola “[...] estimular o desenvolvimento pessoal, social e político do seu aluno, pela ampliação gradativa de suas potencialidades comunicativas” (ANTUNES, 2003, p. 17), é fundamental que, nas aulas de Língua Portuguesa, todas essas práticas sejam estimuladas. Sendo assim, o processo de produção, dentro das aulas, pode ser explorado das mais diversas maneiras, conforme as funções que se pretende cumprir. Segundo a autora, “a escrita varia, na sua forma, em decorrência das diferenças de função que



se propõe cumprir e, conseqüentemente, em decorrência dos diferentes gêneros em que se realiza” (ANTUNES, 2003, p. 48).

Embora não seja o único, provavelmente o gênero mais trabalhado em sala de aula é o da redação escolar. Com relação a esse gênero textual, Ilari acredita que:

Há, mais sutil, o risco de emprestarmos às redações conteúdos que são mais nossos do que dos próprios alunos, ou, ao contrário, de entendermos que uma boa redação é a que resulta de uma maior experiência da vida, a que exemplifica valores mais "elevados" (sejam os critérios de elevação patrióticos, religiosos, cívicos, partidários ou, mais modestamente, práticos). (ILARI, 1978, p. 84).

Ou seja, o autor defende que sejam valorizadas as ideias, a imaginação e a criatividade do redator, em detrimento da forma dos textos e do atendimento às normas gramaticais. Assim, a experiência de vida do estudante teria um papel central no ato de produzir textos. Com isso, a artificialidade do exercício de produção daria espaço à opinião e ao pensamento crítico e argumentativo desse aluno. Dessa forma, Ilari propõe alternativas para a realização das redações no ambiente escolar:

[...] procurei apresentar o exercício de redação como uma oportunidade para explorar a variabilidade da língua: a) quanto às funções a que se presta; b) quanto aos registros em que se emprega; c) quanto às possibilidades que oferece no tocante à informatividade e redundância; além do mais insisti na importância de atentar para aspectos de coesão textual que não são normalmente levados em conta nas gramáticas pedagógicas e nos livros didáticos. (ILARI, 1978, p. 90).

Segundo o autor, dessa forma, a prática da redação escolar englobaria e avaliaria também aspectos linguísticos, além de recursos de coesão e de coerência. Contudo, isso aconteceria de forma coerente com os propósitos do ensino de Língua Portuguesa e de produção de textos em sala de aula. Conforme Evers (2013, p. 55-56):

O texto, nesse sentido, é entendido como um conjunto de *condições* que conduz um leitor e um autor à produção de um evento comunicativo. Enquanto *conjunto de condições*, o texto é um todo de significação e não apenas uma soma de sentenças que estão alinhadas [...].

A redação escolar faz parte dos gêneros textuais denominados por Bakhtin (2003) como gêneros secundários, que são aqueles “[...] normalmente mediados

pela escrita, aparecem em situações comunicativas mais complexas e elaboradas [...]” (SIGNOR, 2008, p. 04). O que diferencia os gêneros primários dos secundários, segundo Bakhtin (2003), é justamente o nível de complexidade em que se apresentam. Assim, a redação escolar se encaixa no gênero secundário justamente porque é produzida no contexto escolar, atendendo solicitação do professor, de maneira mais elaborada.

Sobre os gêneros textuais, os Parâmetros Curriculares Nacionais<sup>7</sup> (PCN) colocam que:

A noção de gênero refere-se, assim, a famílias de textos que compartilham características comuns, embora heterogêneas, como visão geral da ação à qual o texto se articula, tipo de suporte comunicativo, extensão, grau de literariedade, por exemplo, existindo em número quase ilimitado.” (BRASIL, 1998, p. 19).

Essa ideia do ensino de línguas por meio de gêneros tem origem na publicação de documentos oficiais como os PCN, que orientam tais práticas e estratégias por acreditarem que “[...] todo texto se orienta dentro de um gênero e em função das intenções comunicativas [...]” (BRASIL, 1998, p. 21). Tal ideia vai ao encontro do apresentado no início desta seção, por Antunes (2003).

Conforme os PCN:

Os gêneros existem em número quase ilimitado, variando em função da época (epopéia, cartoon), das culturas (haikai, cordel) das finalidades sociais (entreter, informar), de modo que, mesmo que a escola se impusesse a tarefa de tratar de todos, isso não seria possível. Portanto, é preciso priorizar os gêneros que merecerão abordagem mais aprofundada. (BRASIL, 1998, p. 19).

Dessa forma, o documento enfatiza que devem ser priorizados os textos que caracterizam os usos públicos da linguagem, assim como os que aparecem com maior frequência na realidade social e no universo escolar. O objetivo, segundo os PCN, é o de favorecer uma reflexão crítica, bem como o exercício do pensamento,

---

<sup>7</sup> Podemos dizer que os PCN serviram de suporte para a Base Nacional Comum (BNCC), que, atualmente é o documento oficial que orienta e conduz o ensino nas escolas brasileiras, da Educação Infantil até o Ensino Médio. Salientamos, porém, que não utilizamos a BNCC como referência neste trabalho porque não foi discutida, nem implementada nas escolas, nem nas realidades em que lecionamos e que integram este estudo, não estando em vigor nesses locais. Além disso, a BNCC não invalida o que já foi definido pelos PCN, mas, sim, os complementa.

além da fruição estética, a fim de que o indivíduo seja capaz de participação ativa numa sociedade letrada.

Sabemos que há grande discussão em torno dos gêneros discursivos e textuais e dos próprios PCN, porém, acreditamos que não cabe aqui desenvolvê-la, pois não é o foco deste estudo. A restrição é feita em função da utilização de um *corpus* de redações. Entretanto, pela via da descrição do vocabulário manifestado na escrita, podemos contribuir para essa discussão. Nosso trabalho traz somente considerações a respeito da redação escolar como gênero discursivo. Nesse sentido, como o enfoque é léxico-estatístico, centraremos nossas atenções no plano do léxico, pois é nesse contexto que consideramos a palavra e o vocabulário empregados nas redações.

### 2.2.2 Palavra

Conforme Polguère (2018, p. 51), “o termo *palavra* é, pois, de emprego arriscado em Lexicologia”. O autor faz esse comentário porque o termo é realmente muito difícil de definir, pois depende do ponto de vista e do enfoque que será determinado. Biderman (1998b, p. 81) já alertava que “vários são os ângulos sob os quais essa complexa matéria pode ser analisada”.

Os dicionaristas trazem “palavra”<sup>8</sup> como sinônimo de “vocábulo”. Além disso, normalmente misturam seus critérios de definição: sob o critério semântico, seria a “unidade mínima com som e significado, que pode constituir enunciado”; sob o critério escrito, seria a “unidade mínima entre dois espaços em branco, ou entre espaços em branco e sinal de pontuação”; sob o critério oral, seria a “faculdade de expressar ideias, emoções ou experiências; fala”.

Mattoso Câmara propõe uma divisão tripla, defendendo que existem as formas livres, as formas presas e as formas dependentes. Ele coloca que:

Chegamos assim a 3 tipos de formas: 1) forma presa, que só aparece ligada a outra por ela condicionada; 2) forma dependente, que nunca aparece isolada, mas pode aparecer ligada a outra que não é aquela que a condiciona, quando entre ela e sua condicionante se intercalam livremente outras formas; 3) forma livre, que aparece não raro isolada. (CÂMARA, 1967, p. 88).

---

<sup>8</sup> PALAVRA. In: MICHAELIS on-line: Dicionário Brasileiro da Língua Portuguesa. São Paulo: Melhoramentos, 2020. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/palavra/>>. Acesso em: 17 abr. 2020.

Porém, isso não resolve a problemática. Para Biderman (1998b, p. 89), “é a partir da palavra que as entidades da realidade podem ser nomeadas e identificadas. A denominação dessas realidades cria um universo significativo revelado pela linguagem”. Ou seja, é por meio da palavra que o ser humano é capaz de se expressar. E, segundo a autora, essa atividade é específica da espécie humana.

Outra questão levantada por Biderman (1998b) é a não equivalência de palavras entre as línguas. A autora traz exemplos do mandarim e do inglês para relembrar o quanto a falta de equivalência interfere na tradução. Biderman demonstra o problema expondo que, para o tradutor, resolver as questões quando não há correspondência entre as línguas é uma tarefa difícil.

Finatto (2014, p. 171) observa que “o léxico ou o vocabulário, as palavras que, enfim, perfazem uma língua, estão em toda parte e, ao mesmo tempo, associam-se a dimensões diferentes dela”. Sendo assim, cabe ressaltar que o nosso trabalho lida com o ponto de vista estatístico, em que palavra corresponderia a *token*, como usualmente é tratada em Linguística de Corpus. De maneira geral, *token* se refere às ocorrências das palavras num texto, algo que será tratado mais detalhadamente na seção 2.2.5, sobre Linguística de Corpus.

Ao encontro do pensamento estatístico, Polguère (2018) prefere não utilizar o termo “palavra”, empregando, no lugar, os termos: “forma de palavra” e “lexema”. As formas de palavra seriam as variações e flexões possíveis de um lexema, que seria a forma básica ou canônica de uma palavra. Dessa forma, o lexema “ser”, por exemplo, teria as formas de palavra: “sou”, “és”, “foram”, “serão”, etc. Seguindo essa mesma lógica, o lexema “professor” teria as formas de palavra possíveis: “professor”, “professores”, “professora” e “professoras”.

Como vimos, há aspectos que convergem, mas que também divergem entre os diferentes autores tratados aqui. Em função disso, apontamos o nosso conceito, adotado especificamente para este trabalho, na última seção deste capítulo.

### **2.2.3 Léxico/Vocabulário**

O léxico pode ser entendido como o conjunto de palavras de uma língua. Sendo assim, está sempre em mutação, renovando-se constantemente, uma vez

que, “assim como algumas palavras deixam de ser usadas, se tornando arcaísmos na língua, outras novas palavras se incorporam ao vocabulário [...]” (SILVA, 2011, p. 27).

Em relação a essa mutabilidade do léxico, Biderman (1998) afirma que:

Como se sabe, o léxico constitui um sistema aberto de demarcação praticamente impossível. De fato, o crescimento do léxico faz-se numa progressão geométrica, em virtude da criação contínua de palavras novas. O processo das mudanças sociais é um moto-contínuo: a criação humana é incessante e a invenção de novos referentes nunca cessa. (BIDERMAN, 1998, p. 164).

Nesse sentido, podemos caracterizar o léxico como um sistema aberto, capaz de se expandir permanentemente. Conforme Bisognin (2009, p. 27), isso acontece “[...] porque o léxico reflete a dinâmica da língua, que sem se dissociar da sociedade e da cultura, registra a dinâmica da realidade histórica e cultural de uma comunidade linguística”.

Dessa forma, o autor evidencia que as modificações em nosso repertório de palavras disponível para o uso acontecem para espelhar a sociedade, os acontecimentos, a realidade da vida das pessoas. De acordo com Antunes:

A constante expansão do léxico da língua se efetua pela criação de novas palavras (doleiro, internetês), pela incorporação de palavras de outras línguas (deletar, mouse, leiaute [...]), pela atribuição de novos sentidos a palavras já existentes (salvar, fonte, vírus), processos que costumam coexistir e deixar o léxico em um ininterrupto movimento de renovação. (ANTUNES, 2014, p. 31).

Como exemplo, podemos citar o verbo “tweetar”, que já foi acrescentado a dicionários como o Michaelis<sup>9</sup>, por exemplo, sendo grafado como “tuitar” e sendo possível sua conjugação: “eu tuito, tu tuítas, ele tuita”. Essa foi uma das demandas que a modernidade e o uso constante da internet trouxeram, fazendo parte da renovação do nosso léxico, pois “[...] é ao léxico (em sentido amplo, às palavras) que as pessoas recorrem para entender um povo. O léxico é considerado um espelho fiel da vida, refletindo sua organização” (BISOGNIN, 2009, p. 29).

---

<sup>9</sup> TUITAR. In: MICHAELIS on-line: Dicionário Brasileiro da Língua Portuguesa. São Paulo: Melhoramentos, 2020. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/tuitar/>>. Acesso em: 23 jan. 2021.

Nessa mesma direção, Evers (2018, p. 76) afirma que “o fato de serem ‘escolhidas’ determinadas palavras – e não outras – demonstra uma dada realidade vivida, uma vez que as palavras são capazes de testemunhar a história e de sintetizar o pensamento humano”. Ou seja, dentre as opções disponíveis no léxico para o uso, o que é de fato utilizado é o vocabulário que reflete a realidade daquele momento. Por isso, já não utilizamos mais “vosmecê”, por exemplo, e, no seu lugar, usamos “você”: com o passar do tempo, tal palavra foi atualizada. Hoje, pode-se dizer que o uso do “vc” em lugar de “você” talvez seja ainda mais frequente ou mais moderno, inclusive, nos meios digitais e plataformas como *Twitter*, *Facebook*, *Instagram*, *WhatsApp*, etc.

Biderman (1998b, p. 104) expõe que “é preciso lembrar, ainda, que o vocabulário não é criado (ou recriado) pelo indivíduo, mas que ele é adquirido pelo processo social de educação”. Então, as pessoas utilizam de fato aquele vocabulário que faz parte de suas vivências e experiências. Apesar de constantemente serem incorporadas palavras novas à língua, isso não acontece porque um único indivíduo criou e usou diversas vezes uma dada palavra, mas sim por outros motivos, como a necessidade de criação de palavras não existentes em determinado idioma ou de palavras para expressar algo novo, por exemplo. Sobre isso, Estopà (2006, p. 107) declara que:

As palavras podem ser inventadas de diferentes maneiras: uma é criar uma palavra nova, formalmente nova, com as regras que a língua tem (regras que tem desde sempre e que também podem ser novas); outra é dar um novo significado a uma palavra já existente, e outra é tomar palavras de outras línguas (tradução nossa).<sup>10</sup>

Assim, novas palavras são incorporadas ao léxico de uma língua, ou seguindo as regras pré-existentes naquele idioma, ou adquirindo empréstimos de outro, ou recriando significados para formas já existentes na língua. Basílio (2004, p. 09) aponta que:

O léxico, portanto, não é apenas um conjunto de palavras. Como sistema dinâmico, apresenta estruturas a serem utilizadas em sua expansão. Essas

---

<sup>10</sup> Las palabras se pueden inventar de diferentes maneras: una es crear una palabra nueva, formalmente nueva, con las reglas que tiene la lengua (reglas que tiene desde siempre e o que también pueden ser nuevas); otra es dar um nuevo significado a uma palabra ya existente, y otra es tomar palabras de otras lenguas.

estruturas [...] permitem a formação de novas unidades no léxico como um todo e também a aquisição de palavras novas por parte de cada falante.

Dessa forma, o léxico seria algo que vai além de um repertório de palavras, pois inclui também as novas criações, o caráter dinâmico da língua e atualização da realidade. Segundo a mesma autora, o léxico é um “[...] sistema dinâmico, capaz de se expandir à medida que se manifesta a necessidade de novas unidades de designação e construção de enunciados” (BASÍLIO, 2004, p. 09). Nesse sentido está o seu caráter aberto à expansão e à modificação, sempre que necessário. Conforme Bisognin (2009, p. 27), “à medida que ocorrem mudanças na sociedade, a língua a elas se adapta e produz novas unidades léxicas”. O autor afirma que faz parte do desenvolvimento da língua essa criação de novas palavras ou mesmo a reutilização de palavras existentes com diferentes significados. Afinal, a língua é viva, então, está em constante modificação.

Ainda sobre esse tema, Bertonha e Zavaglia (2017, p. 408) mencionam que:

[...] após sua lexicalização [desses novos itens], poderá ou não se realizar o processo de dicionarização e, somente então, passar a compor a nomenclatura de uma obra lexicográfica. No entanto, devemos ressaltar que, a partir do momento que um neologismo é dicionarizado, ele perde a sua natureza de “novo”, passando, desse modo, a fazer parte do acervo lexical daquela língua.

Nesse sentido, é impossível não relacionar o léxico ao dicionário. Embora os dicionários não correspondam ao léxico, podem ser considerados “um espelho” dele, pois fornecem uma imagem do léxico daquela comunidade. Conforme Damim e Peruzzo (2006, p. 94), “para muitas pessoas, o dicionário é apenas um repositório de palavras, com seus respectivos significados, materializado em um único formato”. No entanto, um dicionário é muito mais que isso, pois, assim como colocam as mesmas autoras, assumem papéis pedagógicos, normativos e outros, de acordo com seus propósitos e usuários. Segundo Bertonha e Zavaglia (2017, p. 415),

[...] o dicionário não é uma simples descrição (linguística) nem apenas um convite a explorar suas características reais (social). O dicionário deve ser visto como obra e como fenômeno verbal complexo, como um instrumento de memorização de movimentos culturais, sociais e históricos por meio de itens lexicais repletos de significados e conceitos.

Os mesmos autores ainda registram que é o dicionário que informa e determina qual é o repertório léxico das línguas e, em função disso, exerce papel de autoridade, de obra-referência. Bertonha e Zavaglia (2017) colocam o dicionário como um repositório sócio-histórico-cultural e questionam o porquê dessas mudanças todas na língua e, conseqüentemente, no léxico, em pleno século XXI, ainda demoram tanto para serem dicionarizadas. Segundo eles:

Em pleno século XXI, estimuladas pela mídia e pela informática, sobretudo, pelas redes sociais, a ampliação do léxico ocorre de modo contínuo e acelerado, ao passo que a dicionarização de novas palavras, pelo menos na Lexicografia Brasileira, não segue o mesmo ritmo. Os meios de comunicação, cada vez mais, fazem um exponencial uso do léxico, pois a informação mundial se tornou acessível e compartilhada pela grande maioria das pessoas no planeta, conseqüentemente, o léxico de cada sistema linguístico passou a ser afetado por questões internas e externas a ele. (BERTONHA; ZAVAGLIA, 2017, p. 415).

Em função disso, os autores colocam que há impactos na Lexicografia em todas as partes do globo, oriundos tanto de tecnologia, quanto da comunicação. Nesse sentido, faz-se necessária a atualização destas obras lexicográficas, de maneira que acompanhem o ritmo constante com que o léxico é capaz de se modificar. Conforme Bertonha e Zavaglia (2017, p. 407), “a Lexicografia poderia atentar à recente realidade tecnológica a fim de dicionarizar esses itens lexicais na mesma rapidez com que são criados e veiculados; obviamente, sempre mediante critérios de inclusão justificados teoricamente”.

Ao final de seu texto, os autores concluem indicando que as obras lexicográficas precisam se modernizar, para serem capazes de acompanhar a velocidade do mundo moderno, da tecnologia, e das relações que, por meio dela, se expandem por todo o globo. Encerram afirmando o seguinte: “gostaríamos de acreditar que, num futuro próximo, a Lexicografia brasileira terá condições de ser mais dinâmica que a produtividade lexical [...]” (BERTONHA; ZAVAGLIA, 2017, p. 416).

Ainda em relação às unidades que se constituem como objeto dos dicionários, Barbosa (1995, p. 03) esclarece que:

Quanto aos níveis de atualização e de abstração, parece legítimo estabelecer certas relações entre estes, alguns tipos de obra lexicográfica e as unidades-padrão que constituem seu objeto. Assim, por exemplo, em nível do sistema corresponde a unidade-padrão lexical chamada **lexema**



(Muller); o dicionário de língua tende a reunir o universo dos lexemas, apresentando, para cada um deles, os vocábulos que representam suas diferentes acepções. [...] O vocabulário fundamental, por sua vez, busca reunir os elementos constitutivos de intersecção dos conjuntos vocabulários de uma comunidade, ou de um segmento social, elementos esses que são selecionados pelo duplo critério de alta frequência e distribuição regular entre os sujeitos falantes-ouvintes envolvidos; de maneira geral, o glossário *lato sensu* resulta do levantamento das **palavras-ocorrências** e das acepções que têm um texto manifestado.

Então, para essa autora, as obras lexicográficas têm como objeto o lexema, enquanto o vocábulo remete a cada uma das acepções possíveis de um lexema. E o conjunto desses lexemas reunidos forma um dicionário. Dessa forma, para Barbosa (1995, p. 01), as unidades padrão são: “[...] o lexema, unidade-padrão do universo léxico; o vocábulo, unidade-padrão dos vocabulários técnico-científicos; a palavra, unidade-padrão dos glossários”.

É importante mencionar também que, para Polguère (2018), há uma diferenciação entre léxico e vocabulário. Léxico seria tudo aquilo que temos disponível para uso na língua, enquanto vocabulário seria mais específico, o que de fato é utilizado. Segundo o autor, “o léxico de uma língua é a entidade teórica que corresponde ao conjunto das lexias dessa língua” (POLGUÈRE, 2018, p. 100). Já o vocabulário, por exemplo, pode ser diferenciado entre o vocabulário de um texto e o vocabulário de um indivíduo. O autor coloca que, enquanto “o vocabulário de um texto é o conjunto de lexias utilizadas nesse texto”, “o vocabulário de um indivíduo é o subconjunto do léxico de uma dada língua que contém as lexias dessa língua que o indivíduo em questão domina” (POLGUÈRE, 2018, p. 101). Ou seja, o que um ser humano utiliza para se comunicar é o que ele denomina de vocabulário, o que de fato é usado pelo indivíduo; enquanto o léxico é o conjunto de todas aquelas possibilidades que ele tem disponível. Isso significa que, para o autor, o vocabulário é diferente e varia de um indivíduo para o outro, uma vez que os falantes de uma língua não se comunicam exatamente da mesma maneira, nem têm o mesmo conhecimento de língua.

Diante das considerações expostas acerca do léxico e do vocabulário, ao final do capítulo manifestamos o posicionamento adotado neste trabalho.

## 2.2.4 Palavras Gramaticais e Palavras Lexicais

De acordo com Silva (2011, p. 28-29):

Sabemos que 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. Entretanto, não faremos essa distinção, de maneira que os diferentes itens do vocabulário da língua [...], embora tenham características distintas, fazem parte do léxico aqui considerado.

Então, os diferentes itens do nosso vocabulário, sejam eles gramaticais ou lexicais, serão considerados como integrantes do vocabulário dos estudantes de Ensino Fundamental, público-alvo do nosso trabalho. Porém, se faz necessário diferenciá-los, na medida em que, muitas vezes, trataremos deles separadamente nos estudos piloto e no experimento principal.

Antunes (2014) afirma que a teoria linguística propõe a distinção entre os elementos gramaticais e os elementos lexicais com a finalidade de definir tipos de significação constitutivos das unidades da língua. Assim, a autora coloca que:

- (a) por um lado, as unidades cujo significado remete às coisas, aos fenômenos do mundo da experiência, do mundo extralinguístico, a seus modos de existir, a suas propriedades e manifestações; estabelecem, portanto, a vinculação entre a língua e a experiência; [...]
- (b) e, por outro, as unidades cujo significado remete ao interior do próprio sistema linguístico, constituindo-se em índices da função e da relação que as formas constituintes de um enunciado exprimem; [...] (ANTUNES, 2014, p. 32).

Assim, a autora considera as unidades (a) como lexicais e as unidades (b) como gramaticais (funcionais ou relacionais). Antunes (2014, p. 33) ainda salienta que “[...] essa distinção não implica dicotomia rígida e, como em quase todas as dicotomias relativas à linguagem, carece de rigor e consistência absoluta”. Sobre isso, Estopà (2006, p. 99) complementa, afirmando que:

As palavras se agrupam tradicionalmente em duas grandes classes: palavras lexicais e palavras gramaticais. Esta classificação, entretanto, não pressupõe que as unidades léxicas não tenham propriedades gramaticais, nem que as gramaticais não tenham nenhum peso lexical. Se trata de seu

valor principal. As palavras gramaticais têm uma função primordial de ligação e as lexicais de contribuição semântica. (tradução nossa).<sup>11</sup>

Ou seja, o limiar entre as duas classificações é discutível, na medida em que uma carrega, ainda que não principalmente, características da outra, como é o caso dos pronomes, por exemplo, por apresentarem diferentes tipos. Estopà (2006, p. 100) defende que os pronomes estão entre as duas classificações porque “pertencem a listas fechadas, porém adquirem sentido a cada ocasião que surgem, porque absorvem o sentido da palavra a que se referem” (tradução nossa).<sup>12</sup> Toneli (2010) apresenta o termo “palavras funcionais” e inclui os pronomes pessoais do caso oblíquo entre elas.

As palavras gramaticais (ou funcionais) são aquelas relacionadas ao funcionamento da língua, ou seja, são as que auxiliam as regras e a gramática da língua, cujo seu sentido depende da combinação em que se encontram. Conforme Estopà (2006), formam listas fechadas, pois muito raramente aparecem novas palavras gramaticais. Para a autora, essas palavras seriam as preposições, as conjunções e os determinantes. Já para Antunes (2014, p. 33), que as denomina como “unidades da gramática”, “palavras funcionais” ou “palavras relacionais”, “são apontados como unidades gramaticais: os artigos, os pronomes, as preposições, as conjunções, os advérbios”.

As palavras lexicais, por outro lado, são aquelas com sentido autônomo, independente de onde se encontram e, segundo Evers (2018), que se referem a objetos no mundo. Para Estopà (2006) são os nomes, os adjetivos, os verbos e os advérbios. Nascimento e Isquierdo (2003) as chamam de “palavras plenas” ou “de conteúdo nocional”, considerando os substantivos, os adjetivos e alguns verbos. Antunes (2014) segue perspectiva semelhante, apontando os substantivos, os adjetivos e os verbos como unidades lexicais.

Dessa maneira, numa frase como “O menino brincou com o carrinho novo” temos os dois tipos de palavra possíveis. Os itens “menino”, “brincou”, “carrinho” e

---

<sup>11</sup> Las palabras se agrupan tradicionalmente em dos grandes clases: palabras léxicas y palabras gramaticales. Esta clasificación no presupone, sin embargo, que las unidades léxicas no tengan propiedades gramaticales, ni que las gramaticales no tengan ningún peso léxico. Se trata de su valor más destacado. Las palabras gramaticales hacen una función primera de enlace y las léxicas de aportación semántica.

<sup>12</sup> Pertencem a listas cerradas, pero se llenan de significado em cada ocasión que aparecen porque absorben el significado de la palabra a la cual se refieren.

“novo” são as palavras lexicais, enquanto os itens “o”, “com” e “o” são as palavras gramaticais dessa frase.

Cabe situar que Eugenio Coseriu, linguista e estudioso romeno, diferentemente das perspectivas anteriores, distingue três classes de palavras: as lexemáticas, as categoremáticas e as morfemáticas. Conforme Amorim e Lacerda (2013, p. 183):

As palavras lexemáticas estruturam a realidade extralingüística, as palavras categoremáticas (pronomes) apresentam a forma de estruturação do extralingüístico e as palavras morfemáticas representam as relações com outras palavras. Somente a primeira classe de palavras pertence efetivamente ao léxico e configura enquanto tal o objeto próprio da lexicologia.

Coseriu, então, privilegia o estudo semântico, pois acredita que somente as palavras que classifica como lexemáticas pertencem ao léxico. Tais itens seriam aqueles que se ocupam do significado contido na língua. De acordo com Coseriu (1964b, p. 89) apud Barreiros e Barreiros (2015, p. 401):

As palavras das quais a Lexemática se ocupa são aquelas que desempenham uma função léxica ou uma função social, ou seja, as que estruturam primariamente a experiência. É o que resta depois de se eliminar as determinações gramaticais e categoriais. Portanto, não se considera, para efeito de estruturação, as interjeições, os artigos, os pronomes, as preposições, as conjunções e as partículas de afirmação e de negação além dos nomes próprios e dos numerais. Serão consideradas apenas as palavras lexemáticas, isto é, aquelas que desempenham a “função léxica”, pois representam a “configuração semântica do léxico”.

Em nosso estudo, entretanto, todos os itens encontrados nas produções textuais são considerados parte do vocabulário dos alunos, sendo contabilizados nos resultados estatísticos. De uma perspectiva quantitativa, são consideradas tanto as palavras lexicais quanto as gramaticais para calcular, por exemplo, a variedade vocabular dos textos dos alunos. Em outras vezes, como nosso estudo também é qualitativo, os diferentes itens do vocabulário são separados entre os dois grupos estudados, a fim de analisarmos o perfil vocabular dos estudantes. Trata-se, portanto, de uma análise semântica dos diferentes itens do vocabulário escrito desses alunos, uma vez que buscamos caracterizar esse perfil como um todo.

Até aqui foram expostos os conceitos-chave que são utilizados em nossa pesquisa. Iniciamos pela redação escolar enquanto gênero textual e discursivo, para

que fosse possível, por meio do texto, chegarmos ao nosso objeto principal de estudo, que são as palavras. O conjunto delas, chamado de léxico quando se refere ao todo da língua e chamado de vocabulário quando se refere ao uso específico, é analisado, por vezes diferenciando os itens entre lexicais e gramaticais.

Considerando esse quadro dos principais conceitos teóricos usados em nossa pesquisa, passamos agora à apresentação das áreas de estudo que mais se relacionam com o nosso trabalho, sobretudo a Linguística de Corpus e a Estatística Descritiva Lexical.

### 2.2.5 Linguística de Corpus

Em linhas gerais, a Linguística de Corpus entende a língua como um sistema probabilístico de combinatórias. Esse campo de estudos analisa grandes conjuntos de textos autênticos, os *corpora*, por meio de ferramentas computacionais. O precursor desses estudos no Brasil foi o professor Tony Berber Sardinha, de quem advém muitos dos nossos conceitos. Para o professor:

A Linguística de Corpus ocupa-se da coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas por computador. (BEBER SARDINHA, 2004, p. 03).

Nesse sentido, é a Linguística de Corpus que vai nos auxiliar a descrever o vocabulário escrito das redações dos alunos das duas escolas públicas pesquisadas. Afinal, como salienta Bisognin (2009, p. 62) “é grande a importância do empirismo para a Linguística de Corpus. Isso se deve ao fato de muitos teóricos [...] defenderem que a observação da linguagem real é a única maneira segura de se descrever uma língua”. Este é um dos princípios da Linguística de Corpus: o empirismo, a observação da linguagem real. Essa linguagem, em nosso caso, é a linguagem presente nas redações dos alunos de Ensino Fundamental.

Conforme Silva (2016, p. 122):

A Linguística de Corpus é um ramo da linguística que trata da coleta e análise de materiais escritos ou falados, que são agrupados de acordo com critérios específicos. O corpus por sua vez, é formado de textos de diversos

gêneros, constituindo amostras autênticas da língua que se almeja compreender e pesquisar.

Então, outro dos princípios está na observação e na exploração de textos autênticos, os *corpora*. Um *corpus*, na acepção da Linguística de Corpus, é uma coletânea de dados linguísticos que, na sua composição, seguiu determinados critérios, conforme sintetizados no Quadro 3 a seguir:

Quadro 3 – Critérios para formação de um *corpus*

Característica	Explicação
Origem	Um <i>corpus</i> deve ser autêntico, em linguagem natural. Ou seja, os textos que o compõem não podem ter sido feitos com o propósito de ser objeto de pesquisa, pois, assim, tornam-se artificiais. Nem podem ter sido criados por linguagem artificial, como a de programação, por exemplo. É necessário que sejam escritos por falantes nativos (ou que sejam identificados, quando de aprendizes).
Propósito	A reunião dos textos deve ter como propósito o estudo de fenômenos linguísticos.
Composição	O seu conteúdo deve ser escolhido criteriosamente, segundo o que se pretende pesquisar, levando em conta, principalmente, a naturalidade e a autenticidade.
Formatação	O seu formato deve ser legível por computador, pois seus dados serão processados em ferramentas computacionais.
Representatividade	Um <i>corpus</i> precisa ser representativo daquilo que se pretende observar.
Extensão	Seu tamanho deve ser o mais extenso possível.

Fonte: Adaptado de Berber Sardinha (2004, p. 18-19).

Para Berber Sardinha (2004) muitas são as definições de *corpus*, para inúmeros autores. Contudo, ele considera como a mais completa:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise. (SANCHEZ; CANTOS, 1996, p. 08-09 apud BEBER SARDINHA, 2004, p. 18).

Segundo Beber Sardinha (2004), essa definição seria a mais completa por mencionar todos os critérios já expostos no Quadro 3. Um *corpus*, então, precisa reunir textos autênticos em linguagem natural, pois só assim conseguiremos pesquisar o que de fato ocorre em determinada língua. Além disso, deve ser reunido em formato digital (habilitado, assim, para ser processado em *softwares* e ferramentas computacionais, capazes de prover resultados estatísticos) com o

propósito de estudar fenômenos linguísticos. Dessa forma, os textos são criteriosamente escolhidos, mas o *corpus* em si é artificial, na medida em que é reunido com finalidade de pesquisa, segundo Beber Sardinha (2004). Por fim, um *corpus* precisa ser extenso tanto quanto for possível, na medida em que tal extensão represente a população pesquisada.

Outro dos princípios é a visão probabilística da língua, pois “[...] embora muitos traços linguísticos sejam possíveis teoricamente, eles não ocorrem com a mesma frequência” (BEBER SARDINHA, 2000, p. 350). Nesse sentido, o autor defende que mesmo que o funcionamento da língua permita determinadas ocorrências, algumas são mais prováveis de se realizar do que outras. O exemplo que Beber Sardinha traz é o da frequência dos substantivos, que é muito mais alta do que a das outras classes gramaticais, uma vez que existe número maior de substantivos. Então, conforme Silva (2011, p. 29), “[...] a probabilidade de uma palavra qualquer de um texto ser um substantivo é muito maior do que ela ser de outra classe, já que existem muito mais substantivos na língua”.

Finatto (2009, p. 04) afirma que as palavras preferem algumas associações e rejeitam outras. Dessa forma, se fala em combinações possíveis e combinações prováveis. As combinações possíveis são as que podem acontecer, enquanto as prováveis são as que se realizam com mais frequência, pois são a preferência do usuário da língua. Quanto a isso, Maciel (2002, p. 05) argumenta que:

A recorrência, frequência e estabilidade dessas combinações mostram que elas não são feitas ao acaso, mas são semântica e pragmaticamente restritas. De fato, embora outras combinações léxico-gramaticais sejam teoricamente possíveis, algumas fraseologias e colocações se repetem, enquanto outras são raras ou inexistentes.

Sobre essa estabilidade, Beber Sardinha (2000) afirma que há padrões de regularidade e variações sistemáticas na língua. Assim, a Linguística de Corpus se ocupa da descoberta desses padrões, por meio da análise do uso real da língua.

Outro fator importante é a utilização de ferramentas eletrônicas para o processamento dos *corpora*. Conforme Berber Sardinha (2004), já havia *corpora* e até mesmo Linguística de Corpus antes da existência do computador. Entretanto, como as análises e a própria coleta e composição de um *corpus* eram manuais, os resultados estavam sujeitos ao olho humano. Nesse sentido, poderia haver falhas,

cansaço, imperfeição tanto na composição do *corpus*, quanto na análise dos dados. Além disso, Berber Sardinha (2004) ressalta que essas pesquisas eram focadas, principalmente, no ensino de línguas. Hoje, além de termos a ajuda de *softwares* e ferramentas que tornam o nosso trabalho muito mais rápido e os nossos resultados muito mais precisos, as pesquisas com *corpora* englobam diversas possibilidades.

### 2.2.6 Estatística Lexical nos estudos linguísticos

Pelo fato de o computador existir há mais de meio século, não é possível se falar em “novas tecnologias”, segundo Cúrcio (2006). Pesquisadores utilizam ferramentas eletrônicas, bibliotecas eletrônicas e o correio eletrônico há bastante tempo, cada vez mais e em maior quantidade. Esses recursos vieram para auxiliar e, sobretudo, para otimizar o trabalho humano, de modo que um não substitui o outro, uma vez que os dados gerados pelas ferramentas precisam ser interpretados pelo olho humano. Conforme a autora:

Lembramos também que este processo é constituído por duas partes conjuntas: o trabalho do pesquisador, que é submetido às regras informatizadas, e o programa a ser utilizado, que vai necessitar de diretrizes que serão organizadas tanto pelo programador quanto pelo pesquisador, para que assim o programa possa trabalhar de modo satisfatório. A soma dessas duas partes resulta em uma parceria entre os técnicos que desenvolvem e aprimoram o programa e os pesquisadores da área das ciências humanas que irão utilizá-lo, verificando a melhor maneira de investigar um objeto textual. (CÚRCIO, 2006, p. 02).

Ainda segundo a autora, essa união já resultou em muitas ferramentas e aplicativos, com os quais podemos realizar as mais diversas explorações. A mesma autora explica que:

Ludovic Lebart e André Salem[5] (1994: 15) afirmam que os sucessos obtidos pelas aplicações do método estatístico em muitos domínios, como na física ou na biologia, ganharam espaço nas ciências humanas compreendendo as disciplinas que se utilizam da linguagem, principalmente em estudos de vocabulário. (CÚRCIO, 2006, p. 05).

Nesse sentido, os primeiros trabalhos estatísticos realizados no âmbito das ciências humanas voltaram-se para a estatística descritiva lexical. Cúrcio (2006) salienta que os primeiros trabalhos, anteriores aos métodos quantitativos, foram realizados com o objetivo de comparar o vocabulário de diferentes autores ou de um



mesmo autor durante determinado período. Tais dados quantitativos foram obtidos por meio do uso das ferramentas computacionais. É nesse sentido que os estudos estatísticos se interligam com o Processamento de Linguagem Natural (PLN). Tal união poderia ser muito proveitosa se fosse bem recebida por ambas as partes. Segundo Finatto *et al.* (2015, p. 01):

[...] é preciso dizer que essa integração ou cooperação nem sempre tende a ser bem aceita, sendo recebida com restrições de ambas as partes. Afinal, trata-se de um diálogo entre cientistas da Computação e cientistas da Linguística ainda pouco disseminado no Brasil. Esses cientistas em contato representam, respectivamente, um diálogo entre as assim chamadas Ciências Exatas e as Humanidades.

Assim como a interdisciplinaridade entre humanas e exatas ainda pertença a um cenário futuro ideal, o contato entre o cientista da computação e o linguista também não é algo muito frequente. Conforme as autoras, isso acontece porque o cientista da computação foca na solução por meio de programação, *softwares* e outros; enquanto isso, o linguista encontra falhas nos produtos e sistemas computacionais que lidam com a linguagem humana. Dessa forma, não há integração.

Entretanto, esse diálogo pode proporcionar resultados muito ricos, na medida em que um complementa o outro. Assim como o desenvolvimento de sistemas computacionais facilitam o trabalho com a linguagem, o olhar clínico de um estudioso dela auxilia nas observações e melhorias das ferramentas disponíveis.

O PLN é uma área de investigação da Ciência da Computação que desenvolve sistemas computacionais cujo objeto de pesquisa seja a linguagem. Não tem por objetivo descrever a linguagem (como a Linguística de Corpus, por exemplo), mas sim “[...] criar soluções para problemas bastante pontuais, relacionados com o reconhecimento e a reprodução da linguagem humana em alguma escala” (FINATTO *et al.*, 2015, p. 04). Em virtude disso, lida com comparações, avaliações, abrangência e margem de erro. Segundo Finatto *et al.* (2015), essas margens de erro são esperadas quando os enfoques são lógicos e estatísticos porque as ferramentas computacionais simulam desempenhos humanos. Como exemplo, podemos mencionar o lematizador automático que utilizamos a partir do terceiro estudo piloto, a fim de afinar nossos resultados, sobretudo em relação ao grupo de itens de frequência única – *hapax legomenon*. Tal

lematizador foi criado pelos colegas da Universidade de São Paulo (USP), Cientistas da Computação, e teve margem de erro embutida nos seus resultados. O processo de criação dessa ferramenta e os procedimentos executados por meio desse lematizador são explicados no quinto capítulo deste trabalho, na seção de etapas do procedimento descritivo.

A linguística estatística, por sua vez, “[...] seria uma ramificação da lingüística estrutural, tendo como principal função a descrição estatística do funcionamento das unidades definidas pelo lingüista aos diferentes níveis de análise linguística” (CÚRCIO, 2006, p. 02). Características textuais – como riqueza lexical, abrangência vocabular, atribuição de autoria, entre outros – tornam-se possíveis por meio de uma rápida manipulação informatizada de dados.

Maria Tereza Camargo Biderman introduziu os estudos de linguística estatística no Brasil, em 1967, durante comunicação feita em encontro de linguistas brasileiros. Na sua fala, a autora abordou a Estatística e as Matemáticas como instrumento de pesquisa linguística, os problemas e métodos da Estatística Linguística e as suas aplicações.

Nos seus estudos iniciais, Biderman já considerava que “[...] a língua é um código cujos símbolos obedecem a certas freqüências determinadas e previsíveis” (1967, p. 119), estabelecendo o início desses estudos em nosso país. Na época, ela acrescentava que:

A Estatística Lingüística aplicada ao português viria colaborar eficientemente no aprimoramento das técnicas da lingüística brasileira. Talvez pudéssemos possuir, dentro de certo tempo, calculadoras que nos ajudassem no levantamento dos dados com menor perda de tempo e menor probabilidade de erros. (BIDERMAN, 1967, p. 124).

Anos mais tarde, a autora passou a considerar a face quantitativa da linguagem, afirmando que “muitas teorias foram elaboradas para tratar o fenômeno da linguagem. Uma delas, a Estatística Linguística, considera a face quantitativa da linguagem” (BIDERMAN, 1998, p. 162). Ou seja, a elevada frequência de determinados fenômenos linguísticos constata a sua estabilidade, o que torna possível prever quais fenômenos devem ocorrer preferencialmente, em detrimento de outros. A autora afirma que frequência de letras, fonemas, número de sílabas e

comprimento de palavras, por exemplo, independe da vontade do locutor, pois é condicionado linguisticamente. Em relação a isso, Silva (2011, p. 33) explica que:

Sendo assim, o uso de determinada forma em detrimento da outra pelo falante acontece porque isso independe deste, se relacionando com as formas mais aceitas e usadas pela comunidade dos falantes como um todo, e não pela vontade de cada indivíduo.

Nesse sentido, Biderman (1998) assegura que no uso da linguagem há predeterminações condicionadas linguisticamente, independente do locutor, pois ele não escolhe o que vai usar. Dessa forma, a autora defende que, se analisarmos grande quantidade de dados, conseguiremos prever quais formas são mais frequentes e, portanto, terão preferência no uso, sendo previsíveis de ocorrer.

Assim, Biderman (1998), a partir de um *corpus* de 5 milhões de palavras do português do Brasil escrito (1950-1995) – composto por literatura romanesca, dramática, técnico-científica, jornalística e oratória –, chegou a algumas conclusões sobre o léxico do português brasileiro. Uma delas é a de que os vocábulos gramaticais são usados com maior frequência; outra é a de que 80% de qualquer texto do português é constituído pelas pouco mais de 1000 palavras mais frequentes, que juntas constituem 42% desse *corpus* pesquisado. Essas 1028 palavras mais frequentes são todas as palavras que a autora chama de instrumentais (artigos, preposições, pronomes, conjunções, advérbios e numerais) e algumas palavras que ela chama de plenas (substantivos, adjetivos e verbos).

Nascimento e Isquardo (2003) analisaram 450 redações de vestibular (de 1999 e 2000) de duas universidades do interior do Estado de São Paulo, uma pública e outra particular. Ao contrastarem seus dados com os dados disponíveis do *Dicionário de frequências do léxico do português*, de Biderman (1998), perceberam que havia um grupo de pouco mais de 300 palavras semelhantes entre os dois *corpora*. Isso “[...] indica a existência de um possível núcleo do vocabulário analisado que reúne palavras, provavelmente, usadas em qualquer tipo de texto” (NASCIMENTO; ISQUERDO, 2003, p. 71). As autoras afirmam também que “talvez possam, ainda, por meio de resultados de pesquisas futuras, ser consideradas ‘multiuso’, por aparecerem em qualquer texto” (NASCIMENTO; ISQUERDO, 2003, p. 75).

Biderman (1998) constata que, mesmo na língua escrita, em que, teoricamente, existem mais possibilidades e opções, o vocabulário realmente usado pelo indivíduo não é muito variado. Em função disso, sua pesquisa obteve como resultado 42% de frequência entre as primeiras mil palavras que apareciam como as mais frequentes. Essa frequência de palavras será discutida na seção seguinte, 2.2.7. A lista de ocorrências das palavras segue, com as frequências diminuindo, até chegar nas palavras de única ocorrência, o *hapax legomena*. Tais palavras especificam as características de determinado *corpus* e serão discutidas na seção 2.2.8 deste capítulo.

### 2.2.7 Frequência de palavras

Segundo Berber Sardinha (2004), depois do *corpus* pronto, uma das primeiras e mais fáceis informações que se pode conseguir são as frequências das palavras. Para o autor, essa informação é bastante válida, pois:

Uma das utilidades da análise de frequências de um corpus grande e geral como o Banco do Português é permitir que se determine quais palavras são frequentes e quais são raras. Essa informação não é trivial, pois não se pode obtê-la de outro modo, já que o ser humano não é cognitivamente preparado para armazenar esse tipo de informação. (BERBER SARDINHA, 2004, p. 160).

Com isso, Berber Sardinha (2004) ressalta que a intuição do falante não é capaz de fornecer a frequência exata de uma palavra e nem, portanto, a probabilidade de sua ocorrência. Por isso, é necessária a observação empírica dos textos autênticos, os *corpora*. Conforme o mesmo autor: “A frequência de uso (alta, baixa, intermediária), atributo inseparável da palavra, [...] revela sua ocorrência observada, tem um papel definidor da palavra, fornecendo um traço tão inseparável quanto o sentido [...]” (BERBER SARDINHA, 2004, p. 162). Por outro lado, Coseriu (1987, p. 235) ressaltava que:

A determinação da frequência das palavras tem introduzido, neste campo, um primeiro princípio de racionalização [...], mas estes princípios, ainda que, sem dúvida, importantes, continuam a ser insuficientes para uma aprendizagem racional e sistemática do vocabulário. **A classificação das**

**palavras segundo sua frequência não diz nada a respeito do significado e da designação [...].<sup>13</sup>** (tradução nossa, grifo nosso).

Entretanto, estudos posteriores, como Biderman (1998) e Berber Sardinha (2004), por exemplo, comprovaram que a frequência é importante, pois revela o padrão de uso da língua, auxiliando na sua caracterização, bem como no entendimento da língua como um todo. As frequências são muito usadas na confecção de materiais didáticos para o ensino de línguas, assim como na organização de dicionários e obras lexicográficas. Além disso, comprovam traços de expressão estilística de determinado autor, obra ou período, por exemplo, ou, até mesmo, traços linguísticos de determinado idioma.

Explorar uma língua por meio de uma abordagem que utilize dados nos mostra com precisão o que de fato acontece com essa língua, como ela é usada. Assim, não nos baseamos na intuição ou nos exemplos de livros e de gramáticas, mas, sim, no seu uso real. Dessa forma, as frequências são uma das maneiras de se compreender como determinada língua (ou, em nosso caso, *corpus*) se comporta e quais suas características mais recorrentes.

Biderman (1998) relaciona a frequência das palavras a fenômenos como normas e mudanças linguísticas. De acordo com a autora:

[...] a frequência é uma característica típica da palavra. Aliás, a norma linguística se baseia na frequência dos usos linguísticos. Assim, a norma linguística nada mais é do que a média dos usos frequentes das palavras que são aceitas pelas comunidades dos falantes. E não é só isso. Também as mudanças linguísticas que, no decorrer da história, levam de um estado de língua a outro, advêm das frequências de certos usos em detrimento de outros. (BIDERMAN, 1998, p. 162).

Como podemos perceber, a frequência é relevante para o estudo da língua, na medida em que influencia também o estabelecimento das normas, assim como das mudanças e variações. O fato de determinados traços serem mais recorrentes que outros estabelecem os parâmetros daquela língua.

Como exemplo, Biderman (1998) cita a evolução das línguas românicas, do latim e de seus paradigmas e declinações. Em seguida, relata os estudos em outros

---

<sup>13</sup> La determinación de la frecuencia de las palabras ha introducido en este campo un primer principio de racionalización [...] pero estos principios, aunque, sin duda importantes, siguen siendo insuficientes para una aprendizaje racional y sistemático del vocabulario. La clasificación de las palabras según su frecuencia no dice nada con respecto al significado y a la designación [...].

idiomas como o francês, o espanhol, o italiano e o romeno, que, juntamente com os estudos do português, apresentaram certas características em comum em relação ao vocabulário. A autora destaca que “em todas essas línguas as altas frequências são habitadas pelos vocábulos gramaticais e por um número reduzido de palavras lexicais de significado geral” (BIDERMAN, 1998, p. 163).

Para Berber Sardinha (2004), para que possamos analisar a frequência das palavras, é preciso definir alguns termos que, segundo ele, são chaves nessa discussão:

Quadro 4 – Termos e explicações

<b>Termo</b>	<b>Explicação</b>
<i>Tokens</i>	Palavras, ocorrências ou itens. Cada palavra de uma frase conta como um <i>token</i> , independente dela ser repetida ou não. Exemplo: O(1) meu(2) filho(3) e(4) o(5) meu(6) irmão(7) ganharam(8) o(9) meu(10) presente(11). (11 <i>tokens</i> , conforme numerais nos parênteses).
<i>Types</i>	Formas ou tipos. Número de palavras diferentes. Quando as palavras são repetidas, são contadas apenas uma vez. Exemplo: O(1) meu(2) filho(3) e(4) o(1) meu(2) irmão(5) ganharam(6) o(1) meu(2) presente(7). (7 <i>types</i> , conforme numerais nos parênteses, pois as formas “o” e “meu” se repetem).
Palavras lexicais	Palavras de conteúdo ou plenas, de classe aberta, como os substantivos, os adjetivos e os verbos (algumas classificações consideram alguns pronomes, advérbios, entre outras classes). Exemplo: O meu filho e o meu irmão ganharam o meu presente. (Palavras lexicais: filho, irmão, ganharam, presente).
Palavras gramaticais	Palavras funcionais ou instrumentais, de classe fechada, como os artigos, as preposições, os pronomes, os advérbios, os numerais e as conjunções (algumas classificações consideram alguns pronomes, verbos auxiliares, entre outras classes). Exemplo: O meu filho e o meu irmão ganharam o meu presente. (Palavras gramaticais: o, meu, e).
Lema	Palavras que contêm as suas variações e flexões na sua forma básica ou canônica, chamada lema. Exemplo: O meu filho e o meu irmão ganharam o meu presente. (Lemas: o, meu, filho, e, irmão, ganhar, presente).
<i>Hapax Legomena</i>	Palavras que ocorrem apenas uma vez no <i>corpus</i> , de única frequência. Exemplo: O meu filho e o meu irmão ganharam o meu presente. ( <i>Hapax legomena</i> – considerando a frase como <i>corpus</i> –: filho, irmão, ganharam, presente).

Fonte: Adaptado de Berber Sardinha (2004, p. 165-166).

Ao conhecermos essas nomenclaturas, conseguimos interpretar uma lista de palavras, quando processada por um *software* como o *WordSmith Tools*,<sup>14</sup> usado

<sup>14</sup> *WordSmith Tools*, de autoria de Mike Scott, é um programa de computador muito útil na operacionalização da descrição linguística com *corpus*. É obtido somente pela internet. (BERBER SARDINHA, 2004, p. 83). Disponível em: <<https://lexically.net/wordsmith/>>. Acesso em: 17 abr. 2020.

por Berber Sardinha (2004), por exemplo, ou o *AntConc*,<sup>15</sup> utilizado neste trabalho. Para Berber Sardinha (2004, p. 168), uma lista de palavras

[...] revela um fato conhecido na léxico-estatística e na linguística quantitativa: as frequências de palavras não são distribuídas equitativamente; algumas são mais comuns que as outras. A faixa de frequência mais numerosa (em termos das formas que a compõem) é a de palavras que ocorrem apenas uma vez, os *hapax*.

As frequências das palavras são diferentes porque, seguindo o princípio de Biderman (1998), de que “a frequência é uma característica típica da palavra”, cada palavra tem a sua frequência. Algumas são possíveis de ocorrer, outras são prováveis de ocorrer. Em função disso, suas frequências são diferenciadas.

Conforme já mencionado na seção anterior, Biderman (1998), nos estudos sobre o seu *Dicionário de frequências do português do Brasil*, concluiu que entre as palavras de altíssima frequência estão, sobretudo, as palavras instrumentais e algumas palavras plenas. Essas somam pouco mais de mil palavras que, em porcentagem, chegam a 42% de seu *corpus*. A partir disso, ela revela que “80% de qualquer texto do português é constituído por estas mil palavras, que são reiteradas continuamente” (BIDERMAN, 1998, p. 166). Esse fato já havia sido concluído em pesquisas com *corpora* menores, mais antigos e em outras línguas, fazendo a autora acreditar na hipótese de que “o mesmo deve ocorrer em qualquer língua” (BIDERMAN, 1998, p. 166).

O comportamento dos verbos também foi destacado na sua pesquisa. Biderman (1998) expôs os 20 verbos mais frequentes no português, notando a coincidência deles com pesquisa já realizada em Portugal sobre a língua falada. Nascimento e Isquierdo (2003), ao pesquisarem *corpus* de redações de vestibular, observaram que 13 dos seus 20 verbos mais frequentes são os mesmos da lista de Biderman. Sobre o comportamento desses verbos, Biderman (1998, p. 171) afirma que:

[...] esses vinte verbos mais freqüentes situam-se na escala decrescente de freqüência em posições quase idênticas; isso confirma também que distribucionalmente eles operam de maneira muito similar na língua, não importando o tipo de variáveis lingüísticas consideradas, a saber: língua

<sup>15</sup> *AntConc*, criado por Laurence Anthony, é um *software* de ferramentas de análise de *corpus* para concordância e análise de textos. Disponível em: <<https://www.laurenceanthony.net/software/antconc/>>. Acesso em: 17 abr. 2020.

falada ou escrita [...]. Os resultados demonstram, portanto, que o comportamento lingüístico desses verbos tem-se mantido quase imutável ao longo de duzentos anos. São, pois, verbos muito estáveis no idioma.

A lista dos verbos mais frequentes inicia pelos verbos auxiliares “ser”, “ter” e “estar”, além do verbo “ir”, com valores modais e aspectuais. Biderman (1998) coloca que “estes vinte verbos registram altíssima frequência por serem reiterados continuamente no texto. Dada a centralidade do verbo na articulação do discurso, é normal a enorme repetição dessas palavras” (BIDERMAN, 1998, p. 174).

Outro dado relevante da pesquisa de Nascimento e Isquierdo (2003, p. 75) é que “[...] mais de 85% das [palavras] mais frequentes no *corpus* da pesquisa coincidem com as mais frequentes no *Dicionário de frequências de Biderman*”. Isso colabora com o princípio de Biderman de que algumas palavras seriam consideradas “multiuso” e estariam presentes em todo e qualquer tipo de texto, independente do seu gênero e/ou temática.

A respeito das palavras de baixa frequência, que Biderman (1998) chama de “palavras raras”, discorreremos na seção seguinte, 2.2.8, juntamente com o *hapax legomena*.

### 2.2.8 Curvas de frequência, lematização e papel do *hapax legomena*

Bisognin (2009, p. 119) salienta que

A análise das frequências das palavras é importante para o entendimento da língua como um todo, nos alerta Berber Sardinha (2004). Saberemos a probabilidade de ocorrência de um traço ou estrutura pela observação empírica da frequência de emprego por diversos usuários em determinado contexto.

Nesse sentido, uma lista de frequências de palavras tem sua curva na medida em que progride em ordem decrescente de frequências. Ou seja, a cada nova palavra, a frequência vai diminuindo, até chegar nas palavras de frequência um, chamadas de *hapax legomena*. Conforme Cúrcio (2006, p. 04),

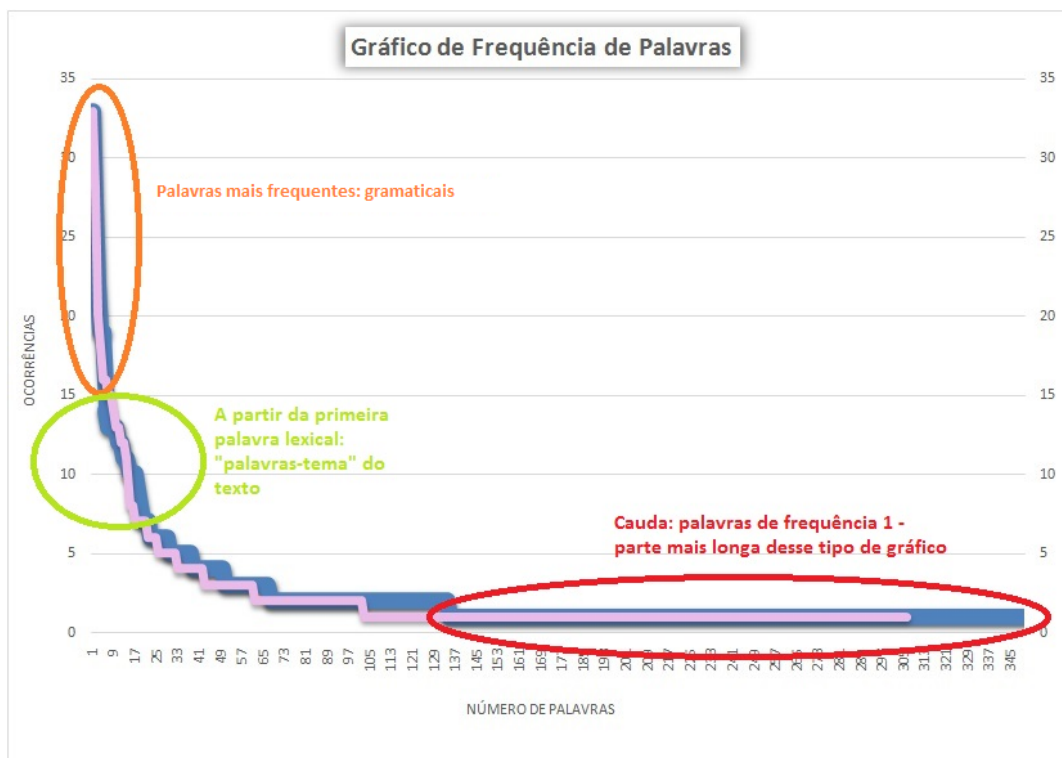
[...] a primeira operação executada pelos programas específicos em análise textual é um corte das seqüências de caracteres em *formas*, de onde eles retiram as *ocorrências*. Por exemplo, se a forma “garrafa” aparece uma vez em um dado corpus, se diz que ela tem uma *ocorrência*, e que tal é a sua *freqüência*. O computador, mais especificamente, o programa é capaz de



distinguir e fornecer uma lista (index ou índice) de todas as formas do texto escolhido, em ordem alfabética ou em ordem decrescente de freqüências.

Quando transportamos esses dados para um gráfico, é possível observar de fato essa diminuição gradativa, que é desenhada sob forma de curva, como no Gráfico 1 a seguir:

Gráfico 1 – Gráfico de frequência de palavras



Fonte: Elaboração própria.

O Gráfico 1 é um gráfico de frequência de palavras que exemplifica o que ocorre com as listas de palavras quando transportadas para uma escala ordenada. Essa ideia de representação por faixas de frequências das palavras em um texto foi formulada na década de 1940, por George Kingsley Zipf, linguista da Universidade de Harvard. Este gráfico, como todos os demais, foi elaborado no programa *Microsoft Office Excel* a partir das listagens de palavras geradas na ferramenta *AntConc*.

O Gráfico 1 inicia com as palavras mais frequentes do *corpus* de estudo analisado que, como já observamos, tendem a ser as palavras gramaticais. Cúrcio (2006, p. 04) lembra que:

Geralmente as formas mais freqüentes na listagem são as chamadas *palavras gramaticais*, também encontradas como *palavras funcionais*[4], como os artigos, os dêiticos, as preposições, os pronomes, as conjunções. Para Bernard, essas palavras podem trazer indicações estilísticas, porém, são menos significantes que os primeiros substantivos listados, pois os mesmos carregam o peso temático.

Em seguida das palavras gramaticais, a partir da primeira palavra lexical, como bem observa Cúrcio (2006), está a segunda camada do gráfico, que é geralmente a de palavras-tópico ou tema do *corpus* e que, por isso, também têm alta frequência. A curvatura desses gráficos tende a ser sempre muito parecida, uma vez que a língua se comporta de maneira semelhante, conforme já discutido anteriormente nos estudos de Biderman (1998), Berber Sardinha (2004) e Nascimento e Isquierdo (2003). Por fim, a última camada, que é também a mais extensa, é conhecida como “cauda” e engloba os itens de frequência um, ou seja, os *hapax legomena*.

Conforme Nascimento e Isquierdo (2003, p. 03):

É evidente que quanto maior e mais diversificado for o vocabulário, tanto maior será o número de unidades diferentes e, em consequência, a frequência média por elas assinalada. É importante ressaltar também que nenhum vocabulário possui uma distribuição homogênea da frequência das palavras, embora seja possível perceber, pela distribuição das suas classes nos dois objetos investigados, uma certa regularidade de estruturas léxicas.

As autoras destacam, no entanto, que essa regularidade não pode ser confundida com homogeneidade, pois nenhum vocabulário tem distribuição homogênea quando se trata das frequências das palavras. Assim, cada *corpus* analisado terá suas especificidades, embora acredite-se que haja regularidades dentro deles, como já foi discutido.

Lebart e Salem (1994) apud Cúrcio (2006) alega que, em algumas situações, pode se “[...] considerar indispensável, antes de todo o tratamento quantitativo sobre o *corpus*, submeter as unidades gráficas oriundas da segmentação automática a uma lematização” (CÚRCIO, 2006, p. 10). Isso foi o que aconteceu em determinado momento de nossa pesquisa. Percebemos que um mesmo item estava sendo contabilizado diversas vezes, em função de suas variações, o que afetava os resultados. Então, constatamos que seria necessário lematizar as listas para que os resultados se tornassem mais precisos.

Lematizar uma palavra é agrupar todas as suas variações ou flexões numa forma básica ou canônica, chamada de lema, conforme já explicado e exemplificado no Quadro 4, da seção 2.2.5 deste capítulo. Segundo Cúrcio (2006, p. 10), lematizar é “[...] dar regras de identificação que permitem reagrupar nas mesmas unidades as formas gráficas que correspondam às diferentes flexões de um mesmo lema”. Desse modo, de acordo com a mesma autora, os verbos ficam no infinitivo, os substantivos e os adjetivos (ainda podemos acrescentar outras formas, como alguns pronomes e artigos, por exemplo) no singular e, quando é possível, no masculino.

Segundo Berber Sardinha (2004, p. 167):

O processo de lematização de *corpora* não é frequente, devido à complexidade da tarefa. A lematização manual é desaconselhável, porque está sujeita a erros, principalmente quando o *corpus* for maior. Feita por computador, ainda assim é sujeita a problemas, já que se serve de algoritmos de *stemming*, ou seja, que operam com base na retirada de partes de palavras [...], um critério não muito seguro.

Sabendo que a tarefa era árdua e que existiam essas duas possibilidades, optamos pela lematização automática. Entendemos que as duas opções trabalham com margem de erro, mas acreditamos que, por ser mais ágil e rápida, a lematização automática seria a mais adequada para o nosso estudo. Embora contenha probabilidade de erro, acreditamos que seu índice de acerto e sua contribuição para os resultados da pesquisa são maiores. Dessa forma, sua utilização favoreceu o nosso estudo.

A lematização foi realizada, então, a partir do terceiro estudo piloto do nosso trabalho, em parceria com os colegas pesquisadores do Núcleo Interinstitucional de Linguística Computacional (NILC-USP). Esse processo, de acordo como foi efetuado com o nosso *corpus* de estudo, será explicado e detalhado no quinto capítulo deste trabalho, na seção de etapas do procedimento descritivo.

Depois que nossas listas foram lematizadas, conseguimos examinar melhor o grupo de itens de frequência única, chamado de *hapax legomenon*. Esse grupo, o maior de uma lista de frequências, conforme podemos visualizar no Gráfico 1, mostra as especificidades e as preferências de determinado autor ou da população ao qual o *corpus* pesquisado representa. Em relação à isso e conforme os resultados de sua pesquisa com o vocabulário escrito das redações de vestibulandos, Nascimento (2001, p. 82) explica que:

Constatamos, portanto, que no grupo de palavras de uma só ocorrência há marcas subjetivas, ora denotando o estilo do autor, ora evidenciando o conhecimento – ou o desconhecimento – da norma lingüística ou do sistema ortográfico. De qualquer forma, são marcas que evidenciam a escolha dos sujeitos no momento da produção do texto.

Biderman (1998, p. 168) afirma que “a esmagadora maioria das palavras raras, *hapax legomena*, são substantivos. Eventualmente ocorrem alguns adjetivos e muito raramente um verbo [...]”. Segundo a autora, isso acontece porque essas palavras só são usadas de fato quando os usuários acionam uma determinada área do conhecimento. Em função disso, são de baixíssima frequência, ocorrendo apenas uma vez no *corpus*. Os verbos, como possuem diversas flexões, costumam se espalhar pelo *corpus*. Quando lematizados, geralmente aparecem no topo da lista de frequências, pois reúnem em seu lema todas as suas formas possíveis.

Nascimento (2001, p. 82) também relata resultados parecidos em seu estudo:

A maioria das palavras desse grupo são as de significação externa, também denominadas plenas. São: substantivos, verbos e adjetivos (nessa ordem hierárquica da frequência) e também um significativo número de advérbios de modo (terminados em *-mente*), que também possuem conteúdo nocional.

Berber Sardinha (2004, p. 169) salienta que “a faixa dos *hapax legomena*, embora seja a mais extensa do ponto de vista das formas, é a menos representativa em termos de ocorrência”. O autor exemplifica, com dados retirados do Banco de Português, que as mais de 180 mil palavras dessa faixa representam apenas 0,15% do todo do *corpus* quando somadas e comparadas às quantidades de ocorrências. Nesse sentido, os resultados de Nascimento (2001, p. 82) também são semelhantes: “destacamos mais uma vez a presença do grande número de *hapax legomena*, que somam mais da metade do total das unidades léxicas, embora a sua frequência corresponda a somente 5,1% do *corpus*”.

O mesmo fato também é atestado em pesquisa com *corpora* de língua inglesa. Recski (2005, p. 253) afirma que “[...] aproximadamente um terço da lista de palavras é constituída por palavras que ocorrem apenas uma vez e que aproximadamente 30 palavras muito frequentes representam um quarto da frequência total de um *corpus*”. O autor comenta as implicações disso para o ensino de uma língua estrangeira, pois, segundo ele “[...] mesmo que os aprendizes leiam

milhões de palavras, muitas delas serão vistas apenas uma vez” (RECSKI, 2005, p. 253).

Biderman (1998) também compara os resultados de sua pesquisa com dados da mesma pesquisa comentada por Recski (2005), denominada *American Heritage Word Frequency Book*. A autora explica que:

Temos aqui outro dado espantoso: estas palavras compõem 25% do *corpus*. Explicando: num total de 42.212 palavras diferentes (lemas) 10.453 palavras ocorreram apenas uma vez. Esse resultado confirma o que havia sido constatado por Richman *et al.* (1971) sobre a língua inglesa na pesquisa intitulada *American Heritage Word Frequency Book*. Aí também os *hapax legomena* constituíram mais de 40% do *corpus*. Creio que no AHWFB não foram eliminados os nomes próprios como em minha pesquisa; essa é provavelmente a razão para se ter atingido uma porcentagem muito maior no inglês. (BIDERMAN, 1998, p. 175).

Em outras palavras, as afirmações de Biderman (1998) corroboram as de Berber Sardinha (2000, p. 344): “em qualquer *corpus*, as formas de frequência um (também conhecidas como ‘*hapax legomena*’) são a maioria”. O pesquisador ainda afirma que:

Baseando-se neste fato, é possível afirmar que o léxico de frequência baixa é o mais comum, isto é, que a maioria das palavras de uma língua é composta de palavras que ocorrem poucas vezes. Em outras palavras, palavras de baixa frequência têm uma probabilidade baixa de ocorrência (1 em 1 milhão, por exemplo) e, já que elas formam a maior parte do vocabulário de uma língua, é necessário usar amostras grandes para que tais palavras possam ocorrer. (BERBER SARDINHA, 2000, p. 344).

Nesse sentido está o nosso interesse particular nessas palavras de frequência um. Por elas serem maioria no vocabulário de uma língua, quando realizada pesquisa léxico-estatística descritiva com *corpora*, é possível que aí estejam as diferenças de vocabulário que apostamos que existem entre os dois grupos estudados. Para descobrir se tal fato ocorre de fato em nosso *corpus* de estudo, analisaremos os itens que compõem a camada *hapax legomena* de nosso *corpus*, contrastando os dados das duas realidades entre si.

### 2.2.9 Riqueza Lexical

A medida tradicionalmente mais utilizada em pesquisas léxico-estatísticas descritivas, sobretudo as que se valem da Linguística de Corpus, chama-se riqueza

lexical, ou variedade vocabular. Ela é calculada por meio de uma razão simples entre número de ocorrências/formas e o número total de palavras. Para lidarmos com a porcentagem dessa medida, multiplica-se por 100 o resultado obtido por meio da razão entre os dois números. Quanto mais próximo dos 100%, maior será a riqueza de determinado vocabulário. Ao mesmo tempo, quanto menor e mais próximo do zero o resultado, menos riqueza há no *corpus* pesquisado.

Sobre isso, Evers (2018, p. 80) explica que:

Essa medida é calculada de forma bastante simples: divide-se o número de palavras diferentes de um texto (os *types* ou tipos) pelo número total de palavras desse mesmo texto (os *tokens*). Assim, se temos um texto com, por exemplo, 87 palavras no total, ou seja, 87 *tokens*, e descobrimos, através de uma análise muito rápida usando ferramentas como AntConc, que 62 palavras nele se repetem (os *types*), então, para calcular a TTR, devemos dividir *types* por *tokens*, ou seja, 62/87. Ao multiplicarmos o resultado por 100, para trabalharmos com uma porcentagem e compreendermos melhor esse número, chegaremos ao resultado de 71,3%, podendo concluir, portanto, que esse texto, de apenas 87 palavras, apresenta uma grande variação vocabular, visto que mais de 70% dele é composto por palavras que não se repetem.

A fim de exemplificar melhor o que Evers (2018) coloca, tomamos como base a mesma frase dos exemplos anteriores, no Quadro 4, da seção 2.2.5 deste capítulo: “O meu filho e o meu irmão ganharam o meu presente”. Nessa frase, conforme já explicado, temos 7 *types* e 11 *tokens*. Ao dividirmos esses números, 7 por 11, temos como resultado 0,63 que, multiplicado por 100, resulta na riqueza lexical de 63%. Então, considerando essa frase de exemplo, a variedade do seu vocabulário é de 63%.

Essa variação, evidentemente, sempre estará ligada ao tamanho de um texto ou *corpus*, pois quanto maior ele for, menor será sua variedade vocabular. Isso é o que afirma Bisognin (2009, p. 166):

Claro está que, quanto maior for esse texto, maior será a possibilidade de muitas palavras serem usadas repetidamente. É o que acabamos de comprovar pela observação dos *corpora*, quanto maior seu número de *tokens*, menor a proporção de *type/token*. É impossível se comunicar utilizando infinito número de palavras sem que sejam, em algum grau, repetidas.

Além disso, conforme Berber Sardinha (2004), quanto maior for o número de *types*, maior será a riqueza e variedade do vocabulário. Quanto menor for o número, mais repetitivo é o universo do vocabulário.

Para que tenhamos um parâmetro, Bisognin (2009) aponta que textos jornalísticos tendem a não repetir palavras, seguindo o preceito da boa redação jornalística. Como exemplo, Belmonte (2007) apud Bisognin (2009), examinou quatro reportagens jornalísticas da área da biodiversidade consideradas de excelência. A riqueza lexical desse *corpus* ficou em torno de 40%. O mesmo autor também analisou textos científicos sobre o mesmo tema e constatou que a riqueza lexical ficou em torno de 29% (BELMONTE, 2007). Dessa forma, é possível perceber que obter resultados altos quando se trata de riqueza lexical não necessariamente atesta a excelência de um texto, pois mais se aproxima de uma lista de supermercado, que cita itens e, ainda assim, pode vir a repetir alguns.

Embora possamos apostar que os textos mais ricos são aqueles que escrevem mais palavras diferentes, resultando numa variedade vocabular maior, as pesquisas nos mostram que o resultado dessa medida não necessariamente se relaciona com a qualidade do texto em questão. Evers (2013, p. 136), ao avaliar textos do exame de proficiência Celpe-Bras, concluiu que “[...] textos mais proficientes – Avançados – possuem menos riqueza lexical do que os textos menos proficientes – Básicos”. Isso foi percebido na sua análise das redações do Concurso Vestibular da Universidade Federal do Rio Grande do Sul (CV-UFRGS) de 2014, em que os textos que obtiveram maiores notas atribuídas pelos corretores não foram aqueles que apresentaram maior variedade de vocabulário (EVERS, 2018).

Finatto, Cremonese e Azeredo (2008), ao analisarem amostra de 30 redações do CV-UFRGS 2008, concluíram que os textos com avaliações mais baixas apresentavam riqueza lexical superior aos do grupo com avaliações mais altas. Segundo as autoras, “esse quadro, em síntese, revela que os textos mais bem sucedidos tendem a lidar com um universo menor de palavras diferentes e que a repetitividade não tende a ser, enfim, uma inimiga do bom texto” (FINATTO; CREMONESE; AZEREDO, 2008, p. 99).

Desse modo, baseadas nesses dados, acreditamos que um texto não necessariamente é melhor, mais bem escrito ou mais eficiente por ter um número mais variado de palavras. Até porque, assim como o que foi relatado nos estudos

apresentados, nos quais, muitas vezes, os escritores utilizavam palavras que não dominavam (seu significado e/ou seu uso), o que ocorre em textos com vocabulário muito variado normalmente é o uso inadequado de determinados vocábulos apenas para não repetir palavras ou com a finalidade de mostrar um vocabulário vasto.

### 2.2.10 *Keywords* e chavicidade

Além da ferramenta *AntConc* (ANTHONY, 2019) que utilizamos para diversas observações em nosso estudo, o *WordSmith Tools* (SCOTT, 2020) é um *software* igualmente importante. De autoria do linguista britânico Mike Scott, foi desenvolvido em 1996, na Universidade de Liverpool, na Inglaterra. É utilizado principalmente por linguistas para trabalhos na área da Linguística de Corpus.

Em nosso estudo, esse sistema foi especialmente útil para a identificação de palavras-chave (*keywords*) do *corpus* de estudo. Conforme Beber Sardinha (2006), um dos instrumentos mais úteis do *WordSmith Tools* é a ferramenta *KeyWords*. De acordo com o autor:

Esse programa permite que o usuário compare uma lista de palavras de seu *corpus* de estudo (ou mais listas de vários *corpora*) com um *corpus* de referência. Desse modo, o usuário obtém uma seleção dos itens lexicais de seu *corpus* de estudo que são estatisticamente mais distintivos. Palavras-chave são, portanto, aquelas cujas freqüências são estatisticamente diferentes no *corpus* de estudo em relação ao *corpus* de referência. (BERBER SARDINHA, 2006, p. 02).

*Keywords*, do inglês palavras-chave, são, portanto, as palavras que apresentam muita diferença de frequência entre dois *corpora*: o de estudo e o de referência. De acordo com o mesmo autor, são necessários dois componentes para que essa análise seja feita. São eles:

- (a) um *corpus* de estudo, representado em uma lista de frequência de palavras. O *corpus* de estudo é aquele que se pretende descrever. A ferramenta *KeyWords* aceita a análise simultânea de mais de um *corpus* de estudo.
- (b) um *corpus* de referência, também formatado como uma lista de frequência de palavras. Também é conhecido como “*corpus* de controle”, e funciona como termo de comparação para a análise. A sua função é a de fornecer uma norma com a qual se fará a comparação das freqüências do *corpus* de estudo. A comparação é feita através de uma prova estatística selecionada pelo usuário (qui-quadrado ou *log-likelihood*). As palavras cujas freqüências no *corpus* de estudo forem significativamente maiores segundo o resultado da prova estatística são consideradas chave, e passam a



compor uma listagem específica de palavras-chave (BERBER SARDINHA, 2004, p. 97).

O programa segue um procedimento para extração dessas palavras-chave, que é diferente da seleção delas pelo humano, que o faria de acordo com a relevância dessas palavras para o texto. Segundo Berber Sardinha (2009, p. 193):

Palavras-chave não são o mesmo que palavras “importantes”. O programa usa um critério estatístico, quantitativo, para identificar as palavras-chave; já uma pessoa usa outros critérios de relevância. Seguindo o mesmo raciocínio, o programa não identifica necessariamente aquelas palavras-chave encontradas, por exemplo, em artigos científicos no campo ‘palavras-chave’ (ou *keywords*, *mot clefs*, etc.).

Nesse sentido reside a importância de um *corpus* de referência, que desempenha papel fundamental nessa análise. De acordo com Berber Sardinha (2009, p. 208), “não há palavras-chave absolutas: elas são sempre relativas ao *corpus* de referência empregado. Duas análises com o mesmo *corpus* de estudo, usando dois *corpora* de referência diferentes, mostrarão resultados distintos”.

Em função disso, nosso estudo buscou utilizar três *corpora* como referência: o Banco de Português, o Lácio Web e o Corpus Brasileiro. Esses *corpora*, que de fato mostraram resultados diferentes quando contrastados com o *corpus* de estudo, são descritos no Capítulo 5, na seção de materiais.

Além disso, o *corpus* de referência precisa ter um tamanho cinco vezes maior que o do *corpus* de estudo e não conter nele esse *corpus*, segundo recomendado por Berber Sardinha (2004). O autor explica o motivo, esclarecendo que:

O *corpus* de referência não deve conter o *corpus* de estudo, porque senão, de certo modo, estaria sendo feita uma comparação do *corpus* de estudo consigo mesmo. Além disso, se o *corpus* de estudo estiver contido no de referência, as frequências do *corpus* “inflacionam” as frequências do *corpus* de referência: devido à soma das frequências, algumas das frequências salientes no *corpus* de estudo podem se obscurecer, e portanto, deixar de indicar palavras-chave. (BERBER SARDINHA, 2009, p. 208).

O autor cita o exemplo da palavra “banana”. Se ela tiver 10 ocorrências no *corpus* de estudo e 1 ocorrência no *corpus* de referência, provavelmente ela será uma das palavras-chave encontradas neste contraste. Porém, se o *corpus* de referência contiver o *corpus* de estudo, os resultados serão diferentes, uma vez que “banana” terá 10 ocorrências no *corpus* de estudo e 11 ocorrências no *corpus* de

contraste. Dessa forma, possivelmente não aparecerá entre as palavras-chave porque sua frequência não será mais destoante entre esses dois *corpora*. De acordo com Novodvorski (2016, p. 177), “tais ressalvas têm o intuito de o *corpus* de estudo não ser filtrado pelo *corpus* de referência, acarretando a perda de traços linguísticos que poderiam ser desconsiderados pelo programa como palavras-chave”.

Os resultados variam de acordo com o padrão escolhido pelo usuário. E, de acordo com Berber Sardinha (2006, p. 03), o tamanho das listas de palavras-chave obtidas como resultado dessa ferramenta é, em geral, problemático, principalmente porque “[...] ultrapassa a capacidade de análise detalhada de que um analista humano dispõe”. Segundo o autor:

O fato de a quantidade de palavras-chave ser em geral maior do que a capacidade de interpretação do analista não se caracteriza como um defeito do programa *KeyWords*. O programa é uma ferramenta, isto é, um instrumento para se dar cabo de uma tarefa; ele não foi desenhado para fazer uma análise inteira, até porque seria impossível, visto que a análise exige interpretação, e os computadores são incapazes de compreender a linguagem e, portanto, de interpretá-la. (BERBER SARDINHA, 2006, p. 03).

Dessa maneira, selecionar parte das palavras-chave para uma análise mais detalhada é um procedimento inevitável, segundo ele. Assim, o olho humano conseguirá analisar, compreender e interpretar os resultados processados pela máquina.

Nesse sentido, Berber Sardinha (2006) aponta para um procedimento de recortes em listas de palavras-chave. Como essas listas normalmente são muito grandes, acaba se tornando um problema para o analista humano explorar todas as palavras. Em função disso, Berber Sardinha (2006) aborda o conceito de palavras-chave exclusivas, que seriam aquelas “[...] que são chave apenas no *corpus* de estudo em questão, em comparação com outros [...]” (BERBER SARDINHA, 2006, p. 02). Isso refinaria a pesquisa, tornando a lista menor e possível de ser explorada pelo analista.

Em nosso trabalho, optamos por estudar somente as palavras-chave de nosso *corpus* de estudo em contraste com os três *corpora* de referência já mencionados. Acreditamos que a análise das palavras-chave exclusivas proporcionaria diversos resultados, originando, assim, um novo estudo. Então, essa nova análise é uma perspectiva de continuidade de nossa pesquisa.

A tabela que o *software* produz como resultado é semelhante a da Figura 2 a seguir. Nessa figura, há diversos itens, dos quais, em nosso trabalho, levaremos em consideração, principalmente, as palavras, as frequências e a chavicidade.

Figura 2 – *Keywords* – modelo

Key word	Freq.	%	RC. Freq.	Rc. %	BIC	P
EU	482	1,82	483.178	0,07	2.199,19	0,0000000000
#	102	0,39	0	0,00	2.055,32	0,0000000000
QUERO	156	0,59	79.846	0,01	901,74	0,0000000000
FILME	141	0,53	100.677	0,01	721,45	0,0000000000
MINHA	157	0,59	155.330	0,02	706,57	0,0000000000
PESSOAS	198	0,75	353.124	0,05	675,14	0,0000000000
ELES	158	0,60	300.044	0,04	516,19	0,0000000000
FAMILIA	45	0,17	1.405	0,00	494,36	0,0000000000
BRASIL	227	0,86	879.688	0,13	460,94	0,0000000000
MAIS	349	1,32	2.326.655	0,33	415,24	0,0000000000
GOONIES	19	0,07	0	0,00	366,29	0,0000000000
NÃO	458	1,73	4.108.713	0,59	359,82	0,0000000000
MEU	106	0,40	187.981	0,03	353,09	0,0000000000
OQUE	18	0,07	0	0,00	345,94	0,0000000000
AMIGOS	67	0,25	44.954	0,01	340,24	0,0000000000
COISAS	76	0,29	85.403	0,01	312,96	0,0000000000
FUTURO	75	0,28	84.666	0,01	307,92	0,0000000000
TAMBEM	27	0,10	615	0,00	305,17	0,0000000000
QUERIA	56	0,21	35.887	0,01	286,05	0,0000000000
BOM	82	0,31	134.837	0,02	280,12	0,0000000000

Fonte: Elaboração própria.

Esclarecendo a Figura 2, na primeira coluna estão as palavras-chave (*Keyword*), depois a frequência no *corpus* de estudo (*Freq.*) e a porcentagem da frequência do item em relação ao total do *corpus* de estudo (%). Seguindo da esquerda para direita, na quarta coluna está a frequência no *corpus* de referência (*RC. Freq.*) e, na quinta, a porcentagem da frequência do item em relação ao total do *corpus* de referência (*Rc. %*). Por fim, nas duas últimas colunas estão a chavicidade (*BIC*) – ou seja, o resultado da estatística de comparação – e o “[...] valor da estatística *p*, que indica o grau de significância da estatística de comparação” (BERBER SARDINHA, 2009, p. 212).

Viana (2010, p. 64) defende que “as principais colunas que dão suporte ao cálculo de palavras-chave correspondem à chavicidade (keyness) e ao valor de  $p$ ”. O autor explica que:

A chavicidade reporta o resultado de um procedimento estatístico pelo qual a ferramenta levanta o quão importante cada palavra-chave positiva é para o *corpus* de pesquisa em relação ao de referência (e vice-versa, no caso das palavras-chave negativas). Quanto maior for o valor apresentado nessa coluna, maior a relevância da palavra em questão.

A coluna  $p$  registra o valor desse índice estatístico. O conceito de  $p$  indica que proporção o resultado encontrado é atribuído ao fator chance. Em outras palavras, quanto menor for o valor registrado para  $p$ , maior é a probabilidade de o resultado realmente expressar uma diferença entre, nesse caso, os dois domínios contrastados. (VIANA, 2010, p. 64).

Ao encontro das ideias de Viana (2010), Berber Sardinha (2009, p. 200) esclarece que “como a lista está ordenada por ordem de chavicidade (‘keyness’), as palavras localizadas no início da lista mostram aquelas cujas frequências são mais discrepantes em relação ao *corpus* de referência”. Em nosso trabalho, interessou-nos, justamente, analisar essas palavras que aparecem no início da lista, que contêm alto índice de chavicidade. Na versão do *software* utilizada em nosso trabalho, esse índice aparece como BIC. Conforme Fromm (2011, p. 12) apud Souza e Rocha (2019, p. 458):

O termo chavicidade, do inglês *Keyness*, é utilizado para estabelecer relação entre o corpus de estudo e um ou mais corpus de referência (em geral, maiores e mais representativos da língua ou tipologia textual em questão). Nesse sentido, demonstra o quanto a palavra (analisada) é representativa em frequência relativa, ou seja, a frequência de seu uso, em determinados corpora é comparada estatisticamente, podendo-se, assim, avaliar se o uso de tal vocábulo é mais ou menos frequente no corpus de análise em comparação com um corpus de referência. Esta comparação estatística pode ser feita por meio de ferramentas do *WordSmith Tools* e programas similares.

Então, a chavicidade é o valor advindo de uma razão estatística oriunda da comparação entre uma lista de frequência de palavras de um *corpus* de estudo e uma lista de frequência de palavras de um *corpus* de referência. Esse valor seria aquele que revela as frequências mais destoantes entre um e outro.

É importante ressaltar que a tabela de resultados gerada pelo *WordSmith Tools* nos mostra não só as palavras com frequência mais alta no *corpus* de estudo,

mas também as palavras com frequência mais alta no *corpus* de referência. Sobre esses dois tipos de palavras-chave, Berber Sardinha (2009, p. 194) explica que:

As palavras-chave podem ser de dois tipos: positivas e negativas. Elas são:  
 Positivas, quando sua frequência é significativamente mais alta no *corpus* de estudo do que no de referência.  
 Negativas, quando sua frequência é significativamente mais alta no *corpus* de referência do que no de estudo.

Em nosso estudo, somente as palavras positivas foram observadas, com o objetivo de auxiliar na caracterização do *corpus* de estudo. As palavras negativas tornaram-se perspectiva de estudo.

Conforme Berber Sardinha (2009, p. 194):

As palavras-chave são úteis na análise linguística para diversos fins, como por exemplo:

- Identificar a temática ('aboutness') de um *corpus* ou de um texto.
- Descrever a organização interna dos textos.
- Localizar marcas indicativas de posicionamento ideológico.
- Traçar um perfil lexical de um autor ou de outros indivíduos.

Dessa forma, observar as palavras-chave do *corpus* de estudo em relação aos *corpora* de referência do português como um todo nos permitiu observar o estilo dos autores, nossos alunos do Ensino Fundamental. Além disso, contribuiu para a descrição e a análise do perfil de vocabulário escrito desse público, que é objetivo deste trabalho.

Magalhães e Novodvorski (2012, p. 296) acreditam que

[...] a análise da chavicidade possibilita, entre outras coisas, a identificação da(s) temática(s) do *corpus* e de indícios de estilo, uma vez que as palavras-chave funcionam como "ponteiros" que indicam áreas no *corpus* que seriam de interesse para o pesquisador.

Dessa forma, acreditamos que a chavicidade realmente apontou para determinados itens do *corpus*, como forma de chamar nossa atenção para algumas características demonstradas por ela.

Ainda, segundo os mesmos autores, "a análise da *chavicidade* (Keyness) está começando a despertar o interesse dos pesquisadores, como uma qualidade textual que daria fortes indícios sobre a temática do texto, **junto a indicadores de estilo**" (MAGALHÃES; NOVODVORSKI, 2012, p. 300, grifo nosso). Nesse sentido,

buscamos a qualidade textual e as marcas de estilo vocabular do *corpus* de estudo, formado pelas redações dos alunos de escolas públicas. Para tanto, analisamos as primeiras 20 *keywords*, de acordo com seu índice de chavicidade, em contraste com os três *corpora* já mencionados. No Capítulo 5 deste trabalho são detalhados os métodos de análise, os *corpora* usados como referência, bem como os resultados obtidos por meio dessa ferramenta.

### 2.3 POSICIONAMENTO DO TRABALHO

Toda a seção 2.2 deste capítulo tratou de revisar a bibliografia das áreas afins ao nosso trabalho, mostrando diferentes perspectivas e pontos de vista, de acordo com a visão de diversos autores. A partir dessa revisão, neste momento, apresentamos o posicionamento adotado nesta pesquisa.

Em primeiro lugar, importa deixar claro que entendemos a redação escolar como um gênero discursivo e textual. É em meio a esse cenário que estamos considerando um perfil de vocabulário que se particulariza. Esse perfil é entendido como um espelhamento do vocabulário desses estudantes via escrita, ainda que a escrita manifestada seja resultado de um processo anterior de leitura (dos textos de apoio, das propostas de produção e/ou dos filmes).

A partir de um *corpus* de redações, nosso foco principal é a descrição do vocabulário escrito conforme a sua apresentação em palavras gráficas. Em meio a esse processo, entendemos léxico de acordo com Antunes (2014). A autora estabelece que “o léxico de uma língua, numa definição mais geral, pode ser visto como o amplo repertório de palavras de uma língua, ou o conjunto de itens à disposição dos falantes para atender às suas necessidades de comunicação” (ANTUNES, 2014, p. 27).

Dessa forma, o léxico é aqui considerado, em um ponto de partida, como uma espécie de “banco de dados” da língua, como todo o repertório de formas à disposição do usuário, sendo um conjunto aberto, capaz de se expandir e de gerar novas unidades. Também, segundo Antunes (2014, p. 29), entendemos que o léxico:

[...] é aberto, inesgotável, constantemente renovável, não apenas porque surgem novas palavras, mas, também, pela dinâmica interna das palavras, que vão e vêm, que desaparecem e reaparecem, que mantêm seus significados ou os mudam, de um lugar para o outro, de um tempo para outro.

Do mesmo modo, de acordo com Polguère (2018), compreendemos o léxico como o vasto conjunto das formas disponíveis para o uso, de maneira mais geral e abrangente.

Já o vocabulário, também acompanhando Polguère (2018), é algo mais restrito e específico, percebido como o conjunto das formas que de fato são utilizadas. O vocabulário é, assim, o conjunto das realizações discursivas das unidades lexicais, que são unidades abstratas (BIDERMAN, 1996, p. 32).

O nosso objeto de estudo, portanto, é o vocabulário escrito dos alunos do Ensino Fundamental de escolas públicas. Esse vocabulário é uma das realizações possíveis do léxico que os estudantes têm à disposição.

Considerando que a língua é constituída pelo plano do léxico e pelo plano da gramática, convém situar a abrangência da nossa análise. Ainda que não estejamos interessadas na sintaxe manifestada nos textos, tampouco em padrões frasais, entendemos que o léxico é um componente muito importante em relação à língua como um todo. Ademais, ao tratar das unidades instrumentais ou gramaticais, estamos incluindo alguns aspectos do funcionamento gramatical. Além disso, consideramos que “[...] léxico e gramática são como que duas faces da mesma realidade, contribuindo de forma complementar para a chamada competência léxico-gramatical dos falantes” (RIO-TORTO, 2006, p. 01).

Naturalmente, em uma pesquisa como a nossa, é necessário um recorte. Além disso, o foco no vocabulário se justifica, dado que a partir dele podemos compreender o léxico e suas correlações.

O léxico espelha quem são as pessoas e as sociedades em que elas se inserem. Também é ele que reflete muitas das mudanças culturais e históricas de uma sociedade, como já mencionado na seção 2.2.3 deste capítulo.

Um exemplo dessas mudanças é o cenário atual em que vivemos, uma pandemia mundial, que incorporou novas formas ao léxico e que ressignificou formas já existentes. Uma situação nova requer novas formas para que sejamos capazes de expressar e de compreender tudo o que está acontecendo. Nesse sentido, palavras como “quarentena”, “pandemia”, “live”, “webinar” ou “lockdown” hoje se tornaram muito usadas por jornais, revistas, televisão, internet e pela população como um todo, incluindo os estudantes do Ensino Fundamental de escolas públicas. Isso mostra o caráter mutável do léxico, que está em constante

alteração e atualização, sendo capaz de se expandir sempre que houver necessidade.

Além disso, o uso das formas estrangeiras acontece, principalmente, porque sua tradução, muitas vezes, não atinge o sentido da língua original por completo. Poderíamos usar as formas “confinamento” ou “isolamento”, entretanto, a população não compreenderia tais sentidos da mesma forma que compreende “*lockdown*”. Nesse caso, sabemos que, muitas vezes, isso acontece porque não há uma explicação clara e objetiva por parte dos líderes e governantes, entretanto, tal discussão não é nosso objetivo aqui. Em função da utilização dessas palavras estrangeiras, elementos originados em outros idiomas vão sendo incorporados ao léxico da nossa língua, algumas vezes aportuguesados como “printar” ou “deletar”, outras vezes mantendo-se o original, como “shopping” ou “skate”.

Ratificando nossa opção pelo léxico e pelo vocabulário, é importante destacar o seu caráter multidimensional, que conecta diferentes planos da língua em funcionamento. A propósito disso, Lorente (2004, p. 20) “[...] apresenta o léxico como um ponto de encontro ou, melhor ainda, uma intersecção de caminhos”. A autora acredita que esta metáfora “ilustra com muita propriedade a ideia de que as perspectivas no estudo do léxico podem ser bem diversas e que esta diversidade não implica necessariamente incompatibilidades ou contradições” (LORENTE, 2004, p. 20). Em relação a isso, Bertonha e Zavaglia (2017, p. 408) complementam, ao afirmar que:

Ao se propor pesquisar o léxico de uma dada língua, o olhar lexicográfico, muitas vezes, irá priorizar um recorte linguístico sob um determinado enfoque teórico. Dessa forma, a depender do embasamento teórico adotado pelo pesquisador, haverá tratamento diferente quanto ao léxico, pois, por um viés lexicográfico, as palavras serão entendidas como unidades lexicais, já por um viés terminológico, serão consideradas como termos, ou por um viés morfológico, serão identificadas por classes de palavras e assim por diante.

Percebemos, então, que são diversos os pontos de vista possíveis para se analisar um mesmo objeto. Em nossa pesquisa, como nosso enfoque é lexicológico, observamos os itens que compõem o léxico de uma língua e que, mais especificamente, compõem o vocabulário escrito desses alunos do Ensino Fundamental.



Nesse sentido, estamos observando as palavras escritas nas redações escolares, a despeito de suas categorias morfossintáticas e processos formativos. A palavra que estamos destacando corresponde a uma unidade de sentido concretizada graficamente.

Diante das várias opiniões sobre a concepção do que seria uma palavra, expostas na seção 2.2.2 deste capítulo, acompanhamos Bisognin (2009, p. 25), para quem palavra é “[...] toda unidade linguística mínima que pode constituir significado, delimitado na escrita por dois espaços em branco e/ou sinal de pontuação”. Sabemos que tal conceito se aproxima da definição de palavra gráfica, e que talvez pareça redutor, mas entendemos que, por utilizarmos ferramentas computacionais, é a noção que mais se adéqua à realidade de nossa pesquisa. Além disso, assim como em Silva (2011, p. 26-27):

[...] aderindo à Linguística de Corpus, neste trabalho cada item gráfico será tomado como um item-palavra, independentemente de sua repetição, variação ou flexões. Isso é o que se denomina, em Linguística de Corpus, de *tokens* (itens, ocorrências) e que corresponde, grosso modo, ao número de palavras gráficas que há num texto.

Ou seja, embora todas as unidades (*tokens*) sejam consideradas e contabilizadas de maneira geral, os diferentes aspectos de uma mesma unidade, tais como valor semântico, também serão levados em conta. Afinal, conforme Evers (2018, p. 75), as análises a partir das quantificações são também qualitativas. Nessa direção, é preciso, por exemplo, distinguir as palavras lexicais das palavras gramaticais.

Segundo Bertonha e Zavaglia (2017, p. 408), “[...] palavras lexicais (vocábulo que têm significado pleno) se apresentam em oposição a palavras gramaticais (aquelas que não têm sentido por si só, pois apontam apenas para dentro da língua, propiciando conexões linguísticas)”. Nesse sentido, neste trabalho, consideramos como palavras lexicais os substantivos, os adjetivos e os verbos.

A despeito dessa escolha metodológica, sabemos que alguns autores questionam o estatuto dos verbos auxiliares, por exemplo, bem como o de alguns tipos de advérbios e de pronomes, considerando-os elementos gramaticais. Contudo, conforme já mencionado, algumas classes de palavras parecem habitar as

duas categorias simultaneamente, pois estão tanto na gramática quanto no léxico da língua. A linha que as separa muitas vezes é tênue.

Ainda assim, optamos por essa escolha na medida em que consideramos as palavras conforme sua classe gramatical, de maneira geral. Por exemplo, o item “não”, em que pese seu sentido autônomo, será considerado como uma palavra gramatical, pois adotamos o critério morfológico. Assim, as palavras consideradas como gramaticais serão os artigos, os numerais, os pronomes, as preposições, as conjunções, as interjeições e os advérbios.

Considerando a necessidade de operacionalizar a descrição do vocabulário em foco, adotamos um critério específico, o que não significa que estamos desconhecendo a dimensão da discussão teórica envolvida. Ainda assim, atribuímos pesos diferenciados aos empregos de elementos lexicais e gramaticais.

É importante salientar que nossa pesquisa está centrada na verificação da distribuição lexical, por unidades, não sendo prioritário o estudo das combinações de palavras, nem dos contextos em que elas ocorrem, como em estudos de prosódia semântica. Entretanto, não há como ponderar a funcionalidade de um dado elemento lexical sem que sejam observadas as suas colocações e construções ao longo de diferentes instâncias de usos. Dessa forma, frisamos que não se trata de apenas observar palavras isoladas em listas de frequência. Ainda assim, como esta pesquisa é uma dissertação de mestrado, foi necessário escolher um foco de estudo.

Sabemos que, ao considerar uma dada palavra “solta”, descontextualizada, é normal que haja dúvidas quanto a sua natureza. Por exemplo, os verbos “pôr” e “parar” e as preposições “por” e “para” só podem ser identificados quando analisados em seus contextos de uso. Entretanto, em um tratamento genérico, optamos por considerar “por” e “para”, independentemente de uso, como itens gramaticais. A decisão se justifica na medida em que não é gerada uma margem de erro significativa em termos estatísticos em função do tipo de texto sob exame, com poucos usos dos verbos “pôr” e “parar”.

Feitas as definições acerca dos conceitos que são utilizados em nosso estudo, cabe também ressaltar algumas escolhas metodológicas associadas às áreas do conhecimento que dialogam com nosso trabalho. Conforme Finatto (2014),

por meio da léxico-estatística podem ser feitas constatações acerca do funcionamento da língua e dos elementos gramaticais. Então:

Importa aqui esclarecer que não se trata de uma estatística lexical igual à que se observa em trabalhos de Processamento de Linguagem Natural (PLN) – que ficam no âmbito da Ciência da Computação, embora muitos procedimentos e técnicas sejam aproveitados. Por outro lado, há uma inter-relação bastante forte dessa Estatística da Linguagem com os princípios, métodos e procedimentos da Linguística de Corpus (LC), tal como hoje a conhecemos, conforme apresentada no Brasil por Berber Sardinha (2004). (FINATTO, 2014, p. 173).

Nessa percepção está o nosso trabalho, pois, conforme já mencionado, utilizamos algumas das ferramentas oriundas do PLN, como um lematizador automático, especialmente desenvolvido para apoiar nossa pesquisa. Contudo, não fazemos uma análise matemática avançada ou multifatorial como, por exemplo, vemos no trabalho de Hartmann (2020). Nossa preocupação é com a descrição pontual e básica dos padrões de vocabulário escrito dos alunos do Ensino Fundamental de escolas públicas. Nossa perspectiva apoia-se nos procedimentos mais elementares da Linguística de Corpus e da Estatística Descritiva Lexical.

Dessa forma, nosso estudo pode ser definido com quanti-qualitativo. Se, em um primeiro momento, utiliza-se dos dados obtidos por meio das ferramentas computacionais, num segundo momento, foi preciso interpretar esses resultados. É muito importante esse olhar do pesquisador sobre os dados, principalmente, em nosso caso, em que o pesquisador conhece o cenário de constituição desses dados, conseguindo, assim, interpretá-los para além dos números. Baker (2004) explicita a relação entre dados estatísticos e análises, mencionando que:

[...] números e frequências são apenas um ponto de partida. Precisamos olhar os dados de maneira mais detida para termos uma compreensão maior dos textos e do que acontece neles, assim como das pessoas que os produziram, de modo a avançar para além do nível básico de descrições e chegar à explanação situada. Valores revelados por números e frequências absolutas servem simplesmente para atrair nossa atenção para alguns aspectos que provavelmente valem a pena serem investigados com mais detalhes. (BAKER, 2004, p. 13, tradução nossa).<sup>16</sup>

---

<sup>16</sup> [...] figures and frequencies are only a starting point. We need to take a closer look at the data and get a feel for the texts and what is happening in them, as well as the people who produce these texts, in order to move beyond low-level description to situated explanation. The value of raw figures and frequencies is simply that they draw our attention to some features that are likely to be worth investigating in more detail.

Acreditamos que os dados probabilísticos e estatísticos muito nos ajudaram, pois não chegaríamos aos mesmos resultados sem o auxílio de uma ferramenta. Entretanto, para que eles façam sentido, necessitam de uma compreensão por parte do analista. Esse olhar humano atento jamais poderá ser substituído pela máquina. Pessoas e máquinas precisam trabalhar juntas em prol de resultados detalhados e acurados. Em relação a isso, Novodvorski e Finatto (2014, p. 15) ponderam que:

[...] a LC associou-se a diferentes aventuras de investigação e praticamente nada rejeitou em termos de parcerias de trabalho – o diálogo tem sido uma marca constante, mesmo com aqueles que encaram a LC apenas como um *modus operandi* computacional e quantitativo. A despeito dessa impressão, claro deve ter ficado nesses, pelo menos, primeiros 10 anos de percurso no Brasil, que vamos muito além de “contar palavras” e que já prestamos uma contribuição muito importante para toda uma comunidade de pesquisa nacional e globalmente conectada. Assim, a aventura tem sido, sim, adequada e, mais do que isso, já muito bem-sucedida.

Nesse sentido, nossa avaliação é quantitativa, pois considera os resultados estatísticos obtidos a partir do processamento de nosso *corpus* por *softwares* e ferramentas computacionais. Todavia, nosso enfoque não é só numérico. Vamos além disso, pois interpretamos, compreendemos e discutimos os dados obtidos como resultados. Não se trata apenas de “contar palavras”, como os autores bem colocam. Até porque, nossa ideia é que esses resultados sirvam como base de propostas de ensino capazes de contribuir com o alargamento do repertório lexical dos estudantes, influenciando assim as suas proficiências leitoras, interpretativas, compreensivas e escritas.

Além disso, como a comparação é também um princípio importante na Linguística de Corpus, nosso estudo também realiza contrastes. Lidamos com dois universos de ensino público: esfera municipal (Porto Alegre) e esfera estadual (Rio Grande do Sul), representados pela produção escrita de seus estudantes. Esses contrastes, além de serem realizados entre os dois públicos em questão, também são feitos com outros materiais, aos quais chamamos de *corpora* de referência e de *corpora* de contraste.

Conforme Berber Sardinha (2004, p. 21), além do *corpus* de estudo, é possível utilizar um *corpus* de referência, entendido como aquele que é usado “[...] para fins de contraste com o *corpus* de estudo”. Em nosso estudo, utilizamos para fins de referência o CorPop (PASQUALINI, 2018), um *corpus* de referência do

português popular escrito do Brasil de 2018. Esse *corpus* foi utilizado como referência de português popular escrito, que seria aquele acessível e compreendido pela maior parte da população. O CorPop foi comparado com nosso *corpus* de estudo, a fim de verificarmos em que medida o vocabulário escrito pelos alunos de Ensino Fundamental está presente nesse tipo de uso do português.

Além do CorPop, utilizamos também, como *corpora* de referência, o Banco de Português, o Lácio Web e o Corpus Brasileiro. Os contrastes com os três *corpora* tiveram como objetivo a análise das palavras-chave, buscando marcas de estilo e especificidades em nosso *corpus* de estudo.

Já para fins de comparação e contraste, utilizamos o que denominamos de *corpora* de contraste, entendidos, em nosso estudo, conforme Tagnin (2013, p. 153), como aquele “[...] que é comparado ao *corpus* de estudo por meio de ferramenta computacional”. Em nosso trabalho, os *corpora* de contraste utilizados foram os textos de apoio às propostas de produção textual (quando existentes, lidos pelos alunos antes da escrita dos seus textos) e textos dos livros didáticos – também trabalhados com esses alunos antes das suas produções (e antes dos textos de apoio). Todos esses materiais, tanto de referência, quanto de contraste, são detalhados e explicados no quinto capítulo deste trabalho, na seção de materiais.

Outra questão que precisa ser ponderada e retomada é o tamanho do nosso *corpus* de estudo. Na seção 2.2.5 deste capítulo, mencionamos que, conforme os critérios e princípios da Linguística de Corpus, um *corpus* precisa ser extenso para que seja representativo. De acordo com Berber Sardinha (2004), quanto maior um *corpus*, mais chances de ocorrerem variados fenômenos, representando, dessa forma, determinada população. Contudo, estudos mais recentes, com *corpora* menores, também têm trazido resultados relevantes, levando os pesquisadores da área a repensarem essa questão, na medida em que se pode considerar que “tamanho não é documento”.

Sobre a representatividade e a extensão de um *corpus*, Pasqualini (2018, p. 41) afirma que:

Entretanto, conforme entendemos, as amostras, a especificidade e o tamanho determinam a representatividade do *corpus* conforme a origem dos textos produzidos e coletados. Consideramos também que conhecer a origem dos textos coletados vai além de simplesmente selecionar textos aleatoriamente da *internet* ou de onde quer que seja.

A afirmação da autora vai ao encontro das especificidades de nosso *corpus* de estudo, uma vez que, enquanto docente dos alunos redatores dos textos, participei do processo de produção das redações que vieram a se tornar o *corpus* principal desta pesquisa. Nesse sentido, e conforme a mesma autora, “participamos inclusive presencialmente do processo de produção de parte do *corpus* que propomos aqui [...]” (PASQUALINI, 2018, p. 41). Cabe ressaltar que todos os textos do *corpus* de estudo foram escritos em sala de aula, na presença da professora e de acordo com as orientações propostas (ANEXO II).

Ainda sobre o critério de extensão do *corpus* como um dos quesitos importantes para sua representatividade, estudos recentes como Lopes *et al.* (2016) e Finatto (2018) questionam a importância desse critério quando comparado, por exemplo, à especificidade e à qualidade dos textos coletados. Segundo Finatto (2018, p. 13), “[...] pudemos validar a ideia de que, conforme o dito popular, ‘tamanho não é documento’, [...] isso também vale para *corpus* e *corpora*”, uma vez que já tivemos ótimos resultados e indicativos com *corpora* menores.

A respeito disso, Pasqualini (2018, p. 42) considera que “[...] um *corpus* pequeno pode ter a mesma validade que um *corpus* grande, desde que consideradas as suas finalidades”. Além disso, Santos (2008) acredita que mais importante que o tamanho é saber o que fazer, como usar e para que um *corpus* é útil.

Nesse sentido, conforme Santos (2008) e Pasqualini (2018), a função merece mais atenção do que o tamanho ou a representatividade de um *corpus*. Até porque “[...] outra questão [...] será a necessidade de criar um novo corpo se não houver nenhum apropriado para as nossas demandas” (SANTOS, 2008, p. 46). No nosso caso, foi preciso configurar uma amostra a partir de um acervo de textos pré-existentes. Reunimos uma amostra balanceada, diante de todas as redações escolares disponíveis, separando-as por ano/ciclo e por escola. Como as temáticas desses textos eram heterogêneas, optamos por homogeneizar o *corpus* com relação à quantidade, às avaliações recebidas e ao tamanho de cada redação. Assim, montamos um *corpus* de estudo estratificado e balanceado. Conforme Biber (2012, p. 14):

A segunda motivação para uma definição prévia da população é que as amostras estratificadas são quase sempre mais representativas do que as

amostras não identificadas (e elas nunca são menos representativas). Isso ocorre porque os estratos identificados podem ser plenamente representados (amostragem de 100%) na proporção desejada, em vez de depender de técnicas de seleção aleatória.

Foram excluídos, por exemplo, textos que não tivessem tratado do tema proposto pela atividade ou que não tivessem um número mínimo de linhas conforme estabelecido. Mais detalhes sobre a seleção dessas redações encontram-se no Capítulo 5, de metodologia.

Defender a validade do tamanho de um *corpus* de estudo em uma pesquisa com redações sempre será algo desafiador frente ao que outros investigadores já produziram. Encontramos pesquisas, de diferentes tipos e naturezas, que já se debruçaram sobre poucos e muitos textos. Um exemplo é a tese de doutorado de Naujorks (2011), que tratou de três textos apresentados ao CV-UFRGS. Por outro lado, há trabalhos que lidam com um universo de mais de mil textos (GRAMA, 2016) e que, ainda assim, podem ser categorizados como pesquisa com *corpus* de pequena dimensão, se considerado o número de *tokens* sob exame. Outros casos a citar são os trabalhos de Evers, que lidaram com o universo de 177 produções textuais (EVERS, 2013) e, depois, com um universo de 341 redações (EVERS, 2018).

De nossa parte, estamos trazendo uma amostra com 154 redações escolares, totalizando 26.424 *tokens*. Tal amostra, conforme acreditamos, é capaz de espelhar o perfil vocabular desses estudantes em questão, principalmente pelos critérios específicos que a compõem, já que o *corpus* de estudo traz textos criteriosamente selecionados entre os disponíveis em nosso acervo pessoal. Dessa maneira, sabemos, por exemplo, quem são esses redatores, quais seus perfis, suas idades e etapas escolares; em quais situações esses textos foram produzidos; quais foram as propostas geradoras e como foram colocadas; quais os textos de apoio foram utilizados. Além disso, as redações dos grupos pesquisados foram por nós cuidadosamente selecionadas para que houvesse semelhança em relação ao número de linhas e à avaliação recebida, de maneira que não houvesse discrepância nesse sentido em relação aos diferentes grupos de alunos. Isso também é detalhado no Capítulo 5, nas seções de materiais e métodos.

Além disso, acreditamos que ao pesquisarmos produções escolares de alunos de Ensino Fundamental, estamos contribuindo com a diminuição de uma

lacuna de pesquisas existentes sobre esse material, mesmo que algum crítico rigoroso possa considerar nosso *corpus* pequeno. De acordo com o que já foi apresentado sobre o estado da arte, neste capítulo, na seção 2.1, grande parte das pesquisas se ocupam de redações de estudantes de Ensino Médio, redações de vestibular, redações do Enem, redações de exames de proficiência, sem haver uma atenção para o texto escrito do estudante do Ensino Fundamental. Em relação aos estudos de vocabulário, encontramos pesquisas com redações de alunos de Ensino Fundamental, mas somente com os mais jovens, crianças que frequentam até o 5º ano. Pouca atenção, em geral, tem sido destinada aos textos dos nossos alunos, de 6º a 9º ano. Assim, nosso trabalho apresenta *corpus* e resultados ainda pouco explorados academicamente, o que justifica a amostra aqui trazida.

Em função do que foi exposto, acreditamos que a qualidade, os resultados obtidos e a representatividade desse *corpus* de estudo superam quaisquer outras questões. Além disso, conforme frisamos desde o início, este estudo tem como propósito a união e o diálogo entre uma experiência docente particular e a pesquisa acadêmica. Então, enquanto docente das realidades abordadas neste estudo, sei que tais produções representam, em boas condições, o perfil de vocabulário escrito desses estudantes.

Por isso é importante a análise das palavras mais frequentes no nosso *corpus* de estudo, comparando os usos dos alunos entre as duas escolas e, até mesmo, entre os dois diferentes anos/ciclos pesquisados. Isso nos possibilitou avaliarmos se havia ou não diferenças de padrões de vocabulário escrito entre os estudantes. Da mesma forma, os contrastes com os *corpora* de referência, bem como com os *corpora* de contraste, também nos permitiram compreender em que medida o vocabulário escrito dos alunos se aproxima de um padrão geral de português escrito e sofre ou não influência dos textos de apoio.

Outro dos itens analisados foi a riqueza lexical, medida estatística que pondera a variedade do vocabulário empregado em um texto ou em um *corpus*. Segundo Berber Sardinha (2009, p. 162-163), essa variedade é chamada *Type-Token Ratio*. Sabemos que essa é uma medida controversa, pois, para o PLN, por exemplo, não é calculada exatamente da mesma forma. Além disso, tal média é também criticada por colocar todas as palavras (gramaticais e lexicais) num mesmo patamar. Entretanto, para nós, linguistas, embora seja um cálculo simples e comum,



trata-se de uma medida produtiva para verificar a variedade vocabular de um dado *corpus*. Em função disso, a utilizamos também para caracterizar nosso *corpus* de estudo, bem como para diferenciar nossos dois públicos.

Tendemos a pensar que o melhor texto é aquele que tem o vocabulário mais variado. Conforme Berber Sardinha (2009, p. 162-163):

Um número maior indica uma variação maior, isto é, há menos repetições de palavras (do mesmo 'type'); um número menor aponta para uma variação menor, pois há mais repetições do mesmo 'type'. Em suma, quanto maior o seu valor, mais palavras diferentes o texto conterá. Em contraposição, um valor baixo indicará um número alto de repetições, o que pode indicar um texto menos 'rico' ou variado do ponto de vista de seu vocabulário.

Contudo, conforme os estudos que já mencionamos no item 2.2.9 deste capítulo, não necessariamente a variedade do vocabulário é um dos itens que faz elevar a avaliação de um texto. Assim como argumenta Finatto (2014, p. 179), “[...] quantitativamente, um texto muito bem escrito não precisa ter, necessariamente, um vocabulário muito variado”. A partir dessas informações é que consideramos tal medida em nosso estudo.

A camada *hapax legomena* também foi um item analisado em nosso estudo. Tal camada é a mais extensa, na medida em que é nela que se encontram os itens de única ocorrência no *corpus*. Conforme Santos (2008, p. 59):

Uma analogia útil para linguística com corpos é o facto de muitas questões raras se tornarem a quase maioria da língua (e não se poderem portanto desprezar): praticamente para qualquer frase que se olhe, ou analise, é possível encontrar algo pouco frequente ou raro... mas, se retirássemos esses casos, ficaríamos sem texto!

A autora resume a importância de se explorar esses itens, pois embora sejam únicos, estão por toda parte e, sem eles, não haveria o texto em si. Em função disso, nós apostamos, desde o início, que as especificidades de cada grupo de alunos-autores seriam importantes para o nosso estudo, na medida em que refletiriam as escolhas particulares dos redatores dos textos. Avaliando a camada *hapax*, pudemos mensurar em que medida haveria uma identidade vocabular dos dois grupos, se suas preferências se aproximam ou se distanciam quando considerado seu vocabulário escrito.

Por fim, também analisamos as palavras-chave do *corpus* de estudo. Palavras-chave seriam aquelas cujas frequências são estatisticamente maiores no *corpus* de estudo do que no *corpus* de referência. Essas palavras funcionam como “ponteiros” que mostram para o pesquisador o que merece ser explorado. As palavras-chave apontam para estilo, temática, posicionamento e perfil do *corpus* estudado, entre outras características. Tal exploração incrementou a nossa descrição do *corpus*, pois nos permitiu enxergar essas questões.

Os itens analisados estatisticamente nos ajudaram a traçar o perfil de vocabulário escrito presente nas redações escolares de nosso *corpus* de estudo. Esse era o objetivo principal do trabalho, que buscou também contrastar as duas esferas de ensino público (municipal e estadual). A ideia, nesse sentido, foi averiguar semelhanças e diferenças entre os dois públicos, a fim de comprovar ou de refutar a hipótese de que existem diferenças de desempenho entre os dois grupos, percebidas subjetivamente pelos professores nas suas atividades docentes.

Como nosso *corpus* de estudo é formado por redações escolares de alunos de Ensino Fundamental, faz-se necessário estabelecer algumas considerações sobre essa etapa da vida escolar. Além disso, acreditamos, baseadas nos estudos a seguir mencionados, que é a partir de um vocabulário mais variado que o indivíduo é capaz de expandir suas habilidades de leitura, compreensão e interpretação, sobretudo os jovens em idade estudantil. Por isso, no próximo capítulo, apresentamos um panorama a respeito do ensino de vocabulário nas aulas de língua materna do Ensino Fundamental.

### 3 O ENSINO DE VOCABULÁRIO NAS AULAS DE LÍNGUA MATERNA NO ENSINO FUNDAMENTAL

Neste capítulo, apresentamos considerações acerca do ensino de vocabulário nas aulas de língua materna do Ensino Fundamental. Primeiramente, o foco é o ensino de vocabulário, uma vez que acreditamos que a diversidade do vocabulário está relacionada às habilidades de leitura, interpretação e compreensão de um indivíduo. Em função disso, consideramos que as aulas de língua materna são um espaço privilegiado para o trabalho com o vocabulário, explorando e expandindo o repertório lexical que o aluno já tem. Por fim, tratamos da etapa específica da vida escolar em que lecionamos e, portanto, temos experiência, além de ser aquela em que estão os redatores dos textos do nosso *corpus* de estudo.

#### 3.1 ENSINO DE VOCABULÁRIO

“O léxico tem um caráter central no processo de ensino-aprendizagem” (GIAMATTEO; ALBANO, 2012, p. 19, tradução nossa),<sup>17</sup> defendem as autoras. É a partir dessa centralidade que entendemos o léxico em nosso trabalho.

Conforme já mencionado, em nosso trabalho, o léxico é entendido como um conjunto de unidades disponíveis para o uso, aberto, que está em constante atualização. Já o vocabulário seria o conjunto mais específico das unidades que de fato são usadas, podendo representar o vocabulário de um texto específico, de uma pessoa, de um determinado autor, época ou período literário, por exemplo. Nesse sentido, nosso objeto de estudo é o vocabulário escrito dos nossos alunos de Ensino Fundamental de escolas públicas do RS.

Nesse contexto, acreditamos que o ensino do léxico desempenha papel importante no ensino de línguas, pois é a partir de um vocabulário mais variado que um indivíduo será capaz de compreender melhor suas leituras. Giamatteo e Albano (2012, p. 19) colocam que:

Quanto ao aspecto comunicativo, não só é importante que ampliem quantitativamente seu vocabulário, como também que adquiram ferramentas para compreender as palavras em diferentes contextos e

---

<sup>17</sup> El léxico tiene un carácter central en el proceso de enseñanza-aprendizaje.

possam usá-las na elaboração de textos coerentes, eficientes e adequados a distintas situações. (tradução nossa).<sup>18</sup>

Sobre o aspecto comunicativo evidenciado pelas autoras, entendemos que o léxico faz parte de qualquer tipo de comunicação, seja ela oral ou escrita, na medida em que, dependendo do que se quer expressar, diferentes palavras são utilizadas. Em relação a isso, Ferreira e Vieira (2013, p. 23) entendem que:

Assim, pode-se considerar que desde as conversas informais entre interlocutores com certo grau de intimidade até os níveis mais elaborados de conversação, há uma forte preocupação com as palavras a serem utilizadas, uma vez que através delas podem-se revelar atitudes proposicionais, ironia, intenções, afetividade, dentre outros.

Dessa forma, o léxico passa a ser considerado em diferentes contextos. Aí reside sua importância e centralidade, enquanto “[...] elemento fundamental para uma adequada proficiência nas práticas de oralidade, leitura e produção de textos [...]” (FERREIRA; VIEIRA, 2013, p. 23). A partir dele, é possível “[...] uma maior compreensão e expressividade nos diversos contextos comunicativos a que os indivíduos estão submetidos/envolvidos” (FERREIRA; VIEIRA, 2013, p. 23).

Conforme Serra (2016, p. 16), “o ensino do vocabulário é importante [...] porque ele dá um conjunto de possibilidades comunicativas e dá acesso a um grande número de outras ideias e de novos mundos os quais o conhecimento deles é de suma importância”. É dessa forma que o aluno, quando exposto a um ensino de vocabulário contextualizado – que lhe mostre as diferentes possibilidades de uso de uma mesma palavra, por exemplo – adquire ferramentas necessárias para articulação de suas ideias, por meio da palavra e de seu repertório vocabular. De acordo com o mesmo autor:

Nesse sentido, o que se entende é que o ensino de vocabulário passa por um trabalho com a escrita, com o texto e com o modo de expressão. Conhecer, adentrar no reino das palavras resulta em um jogo didático interessante e que faz sentido aos alunos. Sem contexto de uso, sem saber “para que serve”, o aluno terá maiores dificuldades para adquirir toda a riqueza que o léxico de sua língua lhe oferece. A prática pedagógica, o conhecimento acadêmico do funcionamento da língua e a didática utilizada pelo professor exercem papel fundamental para que o aluno tenha acesso às possibilidades que seu léxico – sua língua no geral – lhe oferece. (SERRA, 2016, p. 17).

---

<sup>18</sup> En cuanto al aspecto comunicativo, no solo importa que amplíen cuantitativamente su vocabulario, sino que también adquieran herramientas para comprender las palabras em diferentes contextos y puedan usarlas em la elaboración de textos coerentes, eficaces y adecuados a distintas situaciones.

De acordo com esse autor, percebemos que a aprendizagem do léxico também está relacionada à elaboração e à produção de textos, permitindo diversos usos, conforme o contexto, e influenciando em sua coerência. Sobre isso, o autor apresenta o pressuposto que “[...] o conhecimento do vocabulário e seu funcionamento são importantes para a construção de textos adequados e que funcionam em nossa sociedade” (SERRA, 2016, p. 01). O mesmo autor ainda salienta:

Quando se pensa em ensino de vocabulário, não se está pensando em um vocabulário sistemático ou, muito menos, em intermináveis listas de palavras, mas sim, em atividades em que o aluno desenvolva seu vocabulário e, conseqüentemente, sua capacidade tanto da produção quanto de entendimento de textos que circulem em diferentes dimensões e espaços da sociedade. (SERRA, 2016, p. 02).

Ou seja, sabemos que o ensino do léxico, ainda hoje, em muitas escolas, é praticado por meio da repetição, da apresentação de normas gramaticais e das listas de palavras descontextualizadas. Corroborando essa observação, Gil (2016, p. 448) coloca que “o uso da sinonímia, no caso da língua materna, e a tradução literal, na língua estrangeira, são, provavelmente, as práticas mais emblemáticas dessa tradição”. Em relação a isso, Ferreira e Vieira complementam, afirmando que:

Com o passar do tempo, o ensino do léxico continua privilegiando a palavra isolada e sua definição, não exigindo uma reflexão por parte do aluno, limitando-o a identificar apenas um significado para uma determinada palavra empregada em um contexto específico, impedindo que ele perceba as variadas significações da mesma. (FERREIRA; VIEIRA, 2013, p. 20).

E isso é o que percebemos na sala de aula. A visão tradicional da língua e do ensino, que prioriza a gramática e as normas, sobrepondo-as ao uso e à aplicabilidade fazem parte das práticas de ensino mais frequentes. De acordo com Ferreira e Vieira (2013, p. 20):

Tal postura faz com que a maioria dos estudantes apresente tantas dificuldades em lidar com as palavras que chega a comprometer o entendimento do seu texto, principalmente quando tenta utilizar palavras cuja significação não conhece ou mesmo aquelas palavras do seu uso diário, mas que emprega inadequadamente em determinado contexto.

Conforme Gil (2016), essas práticas são chamadas de “atividades descontextualizadas”, pois isolam as palavras de seu contexto, prejudicando a compreensão de seu significado em relação ao todo. Segundo a autora, essas “são práticas que poderiam incrementar o vocabulário do aprendiz se as unidades lexicais das listas, por exemplo, pudessem ser organizadas em campos semânticos, o que tornaria a atividade parcialmente contextualizada” (GIL, 2016, p. 448). Ao organizar uma prática de ensino em redes semânticas, o aprendiz poderia relacionar os sentidos, apreendendo novas formas com mais facilidade.

Tais práticas, inadequadas, precisam ser reinventadas, na medida em que o ensino de vocabulário interligado com outras atividades permite que haja de fato a expansão do repertório lexical do estudante. E isso influencia sua capacidade de compreensão, pois, conforme Serra (2016, p. 02), “muitas vezes, os problemas de leitura e escrita que o aluno enfrenta em seu cotidiano passam pela falta do conhecimento de um vocabulário amplo, além daquele que ele traz de sua comunidade de fala”.

Essa questão foi evidenciada por nós, em sala de aula, e originou muitas das discussões presentes neste trabalho. Percebemos, enquanto docente de Língua Portuguesa de escolas públicas, que o vocabulário dos alunos, na etapa do Ensino Fundamental, em sendo restrito, tende a influenciar suas habilidades interpretativas. Isso é, inclusive, o que considera Serra (2016, p. 16) ao afirmar que “um vocabulário restrito também faz o falante ter sua mente restrita [...]”.

Em função disso, acreditamos que um ensino do léxico diferenciado, contextualizado, interligado a atividades de leitura e de produção pode resultar na ampliação do repertório vocabular que os alunos já têm. E, com isso, pode aumentar também as habilidades de leitura, de interpretação e de compreensão dos estudantes.

Nesse sentido, Serra (2016, p. 02) evidencia que:

Por isso é que o ensino de vocabulário não pode ser desmembrado de uma produção textual ampla e que vise um sentido mais vasto do que simplesmente conhecer novas palavras, mas deve ser um ensino que vise um leitor/ouvinte e um falante/escrevente que saiba adequar sua fala/escrita aos diferentes públicos com o qual ele, como indivíduo social, vivencia em seu cotidiano.

Dessa forma, o ensino do léxico pode contribuir com diferentes aspectos da aprendizagem de um indivíduo. Além de conhecer novas palavras, o aprendiz também será capaz de utilizá-las de forma correta, de adequá-las conforme a situação de uso e de compreendê-las nas suas leituras.

Ferreira e Vieira (2013, p. 23) afirmam que “um dos fatores que crucialmente contribuem para o sucesso na leitura é o conhecimento do vocabulário e nenhuma definição, por mais adequada que seja, iguala a função particularizadora do contexto”. Ou seja:

A prática do ensino de vocabulário é importante e deve figurar nos currículos das escolas, porque o vocabulário não só é um conjunto ou a cristalização das interpretações de como o homem vê o mundo, mas também, um modo como o homem faz e os caminhos pelos quais ele passa para interpretar esse mesmo mundo. (SERRA, 2016, p. 16).

Em suma, diante do exposto, ficou evidenciado que as práticas de ensino de vocabulário ainda persistem no ensino tradicional e descontextualizado, repleto de normas, gramática e repetição. Em função disso, o professor precisa rever tais práticas, a fim de que o aprendiz, ao entrar em contato com o ensino de vocabulário interligado a outras práticas dentro do ensino da língua, seja capaz de expandir seu vocabulário, de compreender seus diferentes usos e seus diversos contextos possíveis, alargando, assim, sua visão de mundo. Além disso, “[...] são escassas as pesquisas sobre o ensino do vocabulário/léxico nas aulas de língua materna.” (FERREIRA; VIEIRA, 2013, p. 19). Dessa forma, o contexto de ensino do léxico precisa ser mais refletido e explorado no âmbito da língua materna, objeto de estudo da próxima seção deste capítulo.

### 3.2 AULAS DE LÍNGUA MATERNA

Conforme nossas pesquisas on-line em bibliotecas ou base de dados como o SciELO, por exemplo, percebemos que poucos são os estudos que tratam do ensino do léxico nas aulas de língua materna. Grande parte das pesquisas que englobam o tema são voltadas para o ensino de língua estrangeira, como garantem Ferreira e Vieira (2013, p. 19): “A partir do trabalho empreendido, ficou evidenciado que o estudo do vocabulário/léxico é notadamente voltado para o ensino de língua estrangeira [...]”.

Dessa forma, nosso estudo pretende contribuir para ampliar essa área, discutindo como o ensino de vocabulário pode e deve ser trabalhado em sala de aula, sobretudo nas aulas de língua portuguesa, nossa língua materna. De acordo com Antunes (2003, p. 122), a meta, o objetivo, a finalidade do ensino de português seria “[...] a ampliação da competência comunicativa do aluno para falar, ouvir, ler e escrever textos fluentes, adequados e socialmente relevantes”. Ao encontro dessas ideias, Serra (2016, p. 01) argumenta que:

Parece não ter chagado às escolas da educação básica do Brasil a fora a ideia de que a aula de língua portuguesa é o espaço em que o aluno desenvolve suas habilidades de leitura e escrita de modo que possa demonstrar, por meio da língua e de outras manifestações da linguagem, seus sentimentos e ideias aos diferentes públicos que compõem a nossa sociedade.

Essa função de ampliar as habilidades comunicativas do aluno, desenvolvendo sua leitura e sua escrita para que ele seja capaz de argumentar, criticar e expor suas ideias é muitas vezes esquecida nas aulas de português. Em vez disso, as regras, a norma, as análises sintáticas e morfológicas, a conjugação de verbos nos mais variados tempos e modos normalmente são sobrepostas à leitura, à interpretação e ao espaço de compreensão e de discussão entre os colegas.

Nesse sentido, Serra (2016, p. 02) coloca que “os diferentes aspectos das aulas, seja o aluno, o professor, o conteúdo, o material didático, a comunidade escolar, tudo deve ser discutido, porque, até onde se pode entender, a falta de habilidade comunicativa é um problema profundo, que tem muitas faces e raízes”. Essa falta de habilidade comunicativa a qual o autor menciona está relacionada às atividades de leitura, interpretação e compreensão dos mais variados textos, geralmente pouco desenvolvidas nas aulas de língua portuguesa e na escola como um todo. Em relação a isso, Serra (2016, p. 02) afirma que:

[...] inúmeros cidadãos que, mesmo em contato todos os dias com os textos escritos das manchetes dos jornais, dos textos da *internet*, ou, ainda texto das placas espalhadas pelas ruas, são completamente incapazes de entender o que aquele conjunto de letras ou desenhos indicam ou estão comunicando.

O autor ainda complementa, alegando que “não muito longe, há aqueles que até conseguem ler, no sentido de decodificar e aproximar o som e a letra, mas que



não passam disso, não vendo lógica alguma naquele texto decodificado, seja qual for sua dimensão” (SERRA, 2016, p. 02). Dessa forma, acreditamos que o trabalho com leitura em sala de aula precisa ser priorizado, na medida em que permite ampliar os horizontes do aprendiz. Essas intersecções com a leitura são discutidas no Capítulo 4 deste trabalho. Além disso, esse trabalho precisa ser relacionado com o trabalho com o léxico e o vocabulário, uma vez que, com um repertório mais amplo, tende a ser mais fácil para o aluno compreender textos, questões e atividades em geral.

Em relação a isso, Ferreira e Vieira (2013, p. 22) acreditam que

Ao buscar um tratamento a ser dado ao léxico no ensino de português, deve-se ter em mente que o encontro de unidades lexicais desconhecidas é uma experiência corriqueira, que deverá se repetir inúmeras vezes pela vida afora depois que o aluno superar a experiência escolar.

Nesse sentido, consideramos que o ensino do léxico deve estar presente ao longo de todo o ano letivo, não como um conteúdo à parte, mas integrado aos outros conteúdos, na medida em que estamos sempre aprendendo palavras novas ou novos significados para formas já existentes. Conforme Ferreira e Vieira (2013, p. 20):

Constata-se com facilidade que na aprendizagem da língua estrangeira, a aquisição do vocabulário é um dos aspectos mais importantes do processo. Por outro lado, na aprendizagem da língua materna, essa aquisição é muitas vezes o único aspecto em que, depois de um certo estágio, o aluno ainda pode progredir. Para Leffa (2000), o domínio da fonologia, morfologia e da sintaxe de uma língua se dá antes de se chegar à adolescência, enquanto o léxico é o único conhecimento que pode ser aumentado, geralmente para o resto da vida, visto que sempre é possível aprender novas palavras.

Dessa forma, as aulas de língua materna devem contribuir para esse aumento do vocabulário que o aluno já tem, pois traz consigo de acordo com suas vivências e experiências anteriores e além da escola. Outro aliado nessa prática é o uso do dicionário. Essas obras precisam ser exploradas nas aulas de português, não somente para buscar o significado, a classe gramatical ou a separação de sílabas de uma palavra, mas, principalmente, para reconhecer os diversos usos possíveis e os contextos em que se pode encontrar um vocábulo. Nesse sentido, Serra (2016, p. 17) argumenta que:

Cumpra lembrar o papel prático da escola nessa discussão: para melhor aquisição por parte dos alunos, a escola deve melhorar suas práticas e seus instrumentos, caso queira mesmo que seu aluno adquira a habilidade de usar e entender os usos linguísticos de nossa sociedade. Juntamente a isso, o dicionário tem sido visto como um novo instrumento didático, importante para o conhecimento dos vocábulos da língua, bem como, do uso desse vocabulário.

Em razão disso, acreditamos que o dicionário é um importante aliado em sala de aula. Tais obras devem auxiliar o trabalho do professor no ensino do léxico, bem como auxiliar o aluno na compreensão e apreensão de diferentes vocábulos.

Nesta seção, discutimos o ensino de léxico nas aulas de língua materna. Acreditamos que ainda há pouca exploração nos contextos escolares brasileiros, algo que precisa ser repensado e revisto, na medida em que adquirimos repertório vocabular ao longo de toda nossa vida, mas ainda é na escola o lugar principal de aprendizagens de um indivíduo, sobretudo o jovem em formação.

Para finalizar, nossas ideias vão ao encontro das ideias das autoras Ferreira e Vieira (2013, p. 31), que acreditam que:

[...] o léxico, entendido como o conjunto de palavras por meio do qual se cumprem os propósitos comunicativos dos seres humanos, deve ocupar um papel de destaque no estudo da língua, pois é por meio dele que se pode integrar linguagem e sociedade, bem como compreender dados ligados à história e à visão de mundo de um determinado grupo de pessoas.

Dessa forma, a escola desempenha papel importante na vida e na formação do ser humano, de maneira que cabe a ela muitas das aprendizagens as quais um indivíduo leva para a vida inteira. Ainda, concordamos novamente com Ferreira e Vieira (2013, p. 31), que consideram que

[...] o ensino do léxico deve ser uma das metas de todos os professores. É dever do educador contribuir para que seus alunos possam se expressar e ver o mundo de uma nova forma mediante um novo vocabulário pois, como se sabe, quanto mais diversificado for o vocabulário de um aprendiz, maior a sua capacidade de se expressar, de compreender e de utilizar corretamente a língua.

Assim, acreditamos que, para ser significativo, todo aprendizado precisa estar interligado, além de conectado com a realidade daquela comunidade aprendiz. Então, as aulas de língua materna são espaço para isso, na medida em que podem trabalhar com a leitura e com a interpretação como forma de ver e de compreender o mundo. O ensino de léxico, em nossa concepção, faz parte disso, pois amplia os

horizontes, mostrando variadas possibilidades e contextos para uma mesma palavra.

A seguir, trazemos algumas considerações acerca desse espaço de ensino de léxico nas aulas de língua materna no Ensino Fundamental. Essa é a etapa da vida escolar na qual lecionamos e trazemos nossa bagagem e experiência, além de ser também a etapa da qual as redações escolares de nosso *corpus* de estudo provêm.

### 3.3 ENSINO FUNDAMENTAL

Atualmente, a educação em nosso país é dividida em educação básica (formada pela Educação Infantil, pelo Ensino Fundamental e pelo Ensino Médio) e educação superior (BRASIL, 1996, p. 08). Dentro do Ensino Fundamental – com duração mínima de oito anos, obrigatório e gratuito na escola pública (BRASIL, 1996, p. 11) –, ainda é possível dividi-lo em dois níveis: Ensino Fundamental I e Ensino Fundamental II. O primeiro nível, chamado de anos iniciais do Ensino Fundamental, compreende do 1º ao 5º ano, e o segundo, chamado de anos finais do Ensino Fundamental, do 6º ao 9º ano.

A partir de 2006, a lei nº 11.274 alterou artigos da LDB de 1996, instituindo a obrigatoriedade do Ensino Fundamental de nove anos. A partir daquele momento, passou a ser obrigatório que toda criança a partir dos seis anos de idade estivesse matriculada na escola (BRASIL, 2006). Tal implementação teve como prazo máximo o ano de 2010.

Nossa prática docente, apesar de já ter passado por outras experiências como Ensino De Jovens e Adultos (EJA) e cursinhos pré-vestibulares, sempre esteve centrada nas séries finais do Ensino Fundamental. Já exercemos a docência em todos os anos que compreendem esse nível, em diversas escolas públicas, tanto administradas pelo Estado do Rio Grande do Sul, quanto por alguns dos municípios da grande Porto Alegre.

Hoje em dia, atuamos inclusive em escola ciclada, cujo modelo educacional, chamado de “Ciclos de Formação”, adotado pela Prefeitura de Porto Alegre a partir do ano de 1996, consiste em uma nova organização de ensino e reorienta o funcionamento da escola. Conforme Mainardes (2010, p. 01):

As origens da concepção de Ciclos de Formação estão relacionadas ao Plano Langevin-Wallon (WALLON, 1977), elaborado em 1946 e 1947, para a reforma da educação francesa no pós-guerra. Nesse Plano, propôs-se que o ensino obrigatório fosse dividido em três ciclos. Os formuladores do Plano defendiam que esse modelo de organização permitiria o desenvolvimento máximo de todos os alunos, independentemente da classe social e dos recursos financeiros das famílias.

O autor explica que a organização escolar, nesse modelo, corresponderia aos ciclos do desenvolvimento humano. Segundo ele, “os nove anos de escolaridade obrigatória são divididos em três ciclos: a) 1º Ciclo (6 a 8 anos de idade); b) 2º Ciclo (9 a 11 anos de idade); c) 3º Ciclo (12 a 14 anos)” (MAINARDES, 2010, p. 02). No Brasil, esse sistema começou a ser implementado a partir de 1995, pelas capitais Porto Alegre e Belo Horizonte.

Nossa pesquisa inclui dados e vivências da segunda metade do Ensino Fundamental, sobretudo no que diz respeito ao 6º e ao 8º anos. Esses anos, do modelo de escola seriado, correspondem, no modelo de educação por ciclos, ao terceiro ano do segundo ciclo, B30, e ao segundo ano do terceiro ciclo, C20, respectivamente. No caso das redações do nosso *corpus*, totalizamos 154 textos. Desses, 60 textos são produções de alunos do 6º ano/B30 e 94 são redações de estudantes do 8º ano/C20.

Conforme os PCN, um dos objetivos do Ensino Fundamental é levar o aluno a “[...] posicionar-se de maneira crítica, responsável e construtiva nas diferentes situações sociais, utilizando o diálogo como forma de mediar conflitos e de tomar decisões coletivas” (BRASIL, 1998, p. 07). Nesse sentido, tal habilidade deve ser trabalhada nas aulas de Língua Portuguesa, visto que são espaços privilegiados para a formação de posicionamento, criticidade, reflexão e opinião do indivíduo.

Serra (2016, p. 16) salienta que “é importante reiterar que o ensino de língua portuguesa, na educação básica, sobretudo com relação à aquisição de vocabulário, ainda tem muito a avançar, principalmente quando se observa o índice de analfabetos funcionais existente no País”.

Nesse contexto, acreditamos ser importante a discussão sobre o ensino do léxico, na medida em que, conforme já mencionado nas seções anteriores, um indivíduo com maior repertório vocabular consegue expressar melhor sua visão de mundo. Dessa forma, consideramos que é na escola que o repertório vocabular que o aluno já tem deve ser explorado, visando à expansão dos seus conhecimentos

lexicais, importantes para sua vida como um todo. Com relação a isso, os PCN enfatizam que:

Essa responsabilidade é tanto maior quanto menor for o grau de letramento das comunidades em que vivem os alunos. Considerando os diferentes níveis de conhecimento prévio, cabe à escola promover sua ampliação de forma que, progressivamente, durante os oito anos do ensino fundamental, cada aluno se torne capaz de interpretar diferentes textos que circulam socialmente, de assumir a palavra e, como cidadão, de produzir textos eficazes nas mais variadas situações. (BRASIL, 1998, p. 19).

Complementando os PCN, Dargel (2011, p. 48) argumenta que:

Partindo do saber do aluno, o professor, por intermédio da prática pedagógica, o levará a conhecer o não-saber. Essa prática pedagógica faz com que professor e aluno enriqueçam, simultaneamente, seu universo. O discurso pedagógico que parte do universo do aluno faz com que ele amplie, indubitavelmente, seu conhecimento lexical.

Ou seja, a prática pedagógica, quando parte do conhecimento prévio e das experiências que o aluno já tem, tende a ser mais eficaz no seu objetivo de que o aluno adquira novos conhecimentos. Dessa forma, o professor também aprende, pois, ao ensinar, enriquece o seu universo. É nesse contexto que se insere o ensino do léxico, permitindo que o aluno amplie seu repertório, quando exposto a novas unidades, ou aos sentidos diferentes de uma mesma palavra em variados contextos. Assim, Ferreira e Vieira (2013, p. 24) afirmam que:

Em relação ao ensino do vocabulário, é válido destacar o trabalho de Bezerra (2000) que considera que, apesar da presença constante do vocabulário na vida cotidiana das pessoas, o seu estudo nas aulas de português do ensino fundamental e médio ainda acontece de forma descontextualizada, pois não se considera o texto como um todo e menos ainda a relação entre o oral e o escrito em registros variados.

Conforme já mencionado nas seções anteriores, as práticas escolares quanto ao ensino do léxico seguem sendo principalmente por meio da repetição, das listas de palavras descontextualizadas, ou, segundo Dias (2006), por meio de exercícios dos livros didáticos que utilizam somente antonímia e sinonímia. A autora menciona que:

Temos notado, ao longo da nossa experiência no Ensino Fundamental, que a escola não dá a devida atenção ao trabalho de ampliação do vocabulário dos alunos. Na verdade, a maioria dos professores pouca atenção tem dado

aos problemas concernentes ao léxico. Além de estar sempre preocupada com o conteúdo programático, a maioria não tem formação nas Áreas de Lexicologia e Lexicografia, não percebendo, por isso, o papel do léxico no mundo comunicacional em que o aluno está inserido. (DIAS, 2006, p. 195).

Esta é outra questão importante: como o ensino do léxico não constitui conteúdo específico, como substantivos, análise sintática interna ou crase, por exemplo, seu ensino não está interligado ao conteúdo programático mínimo de um ano letivo. Dessa forma, muitas vezes, os professores priorizam o cumprimento daqueles conteúdos cujo ensino está programado para aquele determinado ano, esquecendo de entrelaçar o léxico como parte dos outros conteúdos gramaticais e das análises e interpretações de textos. Em consonância com Dargel (2011, p. 16), acreditamos que:

Nesse sentido, constatamos, pois, a necessidade de o ensino de Língua Portuguesa valorizar o ensino de vocabulário, propiciar, ao aluno, o entendimento do valor da palavra no texto e o conhecimento de suas diversas possibilidades de sentido. Desse modo, o aluno será capaz de compreender textos de estrutura e gêneros variados, ao aproveitar o conhecimento lexical trazido de casa e do convívio com as pessoas de maneira geral, desenvolvê-lo e acrescentá-lo ao adquirido na escola.

É nas séries finais do Ensino Fundamental que o aluno passa a estudar as disciplinas separadamente, com diversos professores, especializados em cada área. Então, entendemos que também é a partir dessa fase que o ensino do léxico precisa ser desenvolvido com mais atenção com os educandos. Como vimos, o léxico faz parte das nossas vidas desde que aprendemos a nos comunicar, seja por meio da fala ou da escrita. Mas é na escola que esse repertório precisa ser estimulado e desenvolvido, a fim de expandir os conhecimentos prévios do aprendiz. Essa expansão do repertório vocabular vai ajudar o aluno nas suas atividades escolares de leitura e compreensão em todas as disciplinas. Além disso, vai auxiliá-lo por toda sua vida, na medida em que sempre precisamos interpretar e expressar nossas opiniões e visões de mundo. E isso se faz por meio da palavra. Quando compreendemos uma palavra no seu contexto e passamos a utilizá-la em nosso dia a dia, de fato ampliamos nosso repertório vocabular.

Considerando a importância do assunto, queremos contribuir com os estudos da área, abordando o perfil de vocabulário escrito do nosso aluno do Ensino Fundamental. Afinal, conforme Biderman (1998a, p. 179):

Verifica-se, portanto, que esse tipo de pesquisa no âmbito do vocabulário é de particular interesse para o ensino, pois somente com o conhecimento sobre quais são os vocábulos que merecem maior atenção, ou sobre quais são os fatores que desencadeiam semelhanças e/ou dessemelhanças com relação ao uso, é que se pode “evitar o empirismo na escolha do vocabulário para fins didáticos.”

No próximo capítulo, exploramos as intersecções entre o vocabulário e a leitura. Se a leitura nos faz enxergar o mundo, é por meio da palavra escrita que ela existe. Dessa forma, tratamos da ligação entre leitura e escrita, sempre tendo em mente que nosso *corpus* de estudo, as redações escolares, foram resultantes de atividades de leitura.

## 4 O VOCABULÁRIO NA ESCRITA E INTERSECÇÕES COM A LEITURA

Sabemos que a escrita é o resultado da leitura, assim como, muitas vezes, a leitura também pode ser o resultado da escrita. Em função dessa ligação entre as duas, este capítulo mostra quais são as intersecções entre esses dois processos, além da relação deles com o vocabulário, eixo principal de nosso estudo.

Em relação a essa intersecção, Antunes (2003, p. 67), afirma que:

A atividade da leitura completa a atividade da produção escrita. É, por isso, uma atividade de interação entre sujeitos e supõe muito mais que a simples decodificação dos sinais gráficos. O leitor, como um dos sujeitos da interação, atua participativamente, buscando recuperar, buscando interpretar e compreender o conteúdo e as intenções pretendidas pelo autor.

Conforme a complementaridade apontada por Antunes é que entendemos nosso *corpus* de estudo. As redações escolares que o constituem foram produzidas com base em atividades anteriores de leitura e de interpretação, seja de textos (que utilizamos, na pesquisa, como *corpora* de contraste), seja de filmes. Dessa forma, a atividade de escrita completou uma atividade anterior, a de leitura, resultando na produção de um gênero típico da sala aula, conforme já discutido no capítulo anterior: a redação escolar.

Com relação à leitura, a escola é normalmente o contexto em que as crianças entram em contato com ela. Rodrigues (2014) acrescenta que a escola é também o local onde a maioria dos alunos aprende a ler. Sendo assim, o gosto de um indivíduo pela leitura depende, geralmente, de como acontecem esses primeiros momentos.

A importância da leitura é inegável. Acreditamos que é essencial que saibamos ler e compreender o que acontece ao nosso redor, em nossa sociedade e no mundo, para que consigamos sobreviver como cidadãos, exigindo nossos direitos e cumprindo nossos deveres. Segundo Rodrigues (2014, p. 12):

Sem a competência da leitura, dificilmente o indivíduo sobreviverá numa sociedade que é implacável com os analfabetos e iletrados. Falamos em competência, pois consideramos que é o saber em ação; não basta saber ler, é necessário perceber o que lemos, aprender os seus significados, inferir os sentidos ocultos, para posteriormente mobilizar esses conhecimentos noutras áreas ou noutras atividades, sem esquecer o papel preponderante que a disciplina de língua materna tem nesta aprendizagem.



Nesse sentido, Rodrigues traz à tona a discussão de que ensinar a ler e a escrever são um compromisso de todas as áreas, não se restringindo ao cenário do ensino de Língua Portuguesa. O trabalho com a língua materna tem e deve ter papel fundamental no desenvolvimento dessas proficiências. Como todas as outras disciplinas escolares também precisam que um aluno seja capaz de ler e de escrever, é imprescindível que também essas disciplinas auxiliem no progresso dessas capacidades. Em relação a isso, Rodrigues (2014, p. 12) acrescenta que:

Reflexo dessa consequência é o facto de existirem muitos colegas de diferentes áreas curriculares que se queixam dos alunos que “não sabem ler os enunciados”; “não respondem corretamente às questões, porque não leem os itens até ao fim”; “não sabem o que quer dizer a palavra x”, entre outras lamentações, o que nos leva a depreender que a leitura e o conhecimento do vocabulário são essenciais para o sucesso educativo em todas as áreas do saber.

Nossas ideias vão ao encontro das ideias da autora, uma vez que tais saberes são os que um aprendiz levará para toda sua vida. A autora ainda complementa, afirmando que “[...] o entendimento daquilo que se lê e a habilidade na leitura terão reflexos no rendimento escolar dos alunos em todas as áreas do saber e na sua vida profissional” (RODRIGUES, 2014, p. 30-31). Nesse sentido está a importância da leitura, pois ela reside em todas as áreas e permanecerá sendo necessária a todo o momento. Em consonância a esse pensamento, Guedes e Souza (2011, p. 03) declaram que:

Ler e escrever são tarefas da escola, questões para todas as áreas, uma vez que são habilidades indispensáveis para a formação de um estudante, que é responsabilidade da escola. Ensinar é dar condições ao aluno para que ele se aproprie do conhecimento historicamente construído e se insira nessa construção como produtor de conhecimento. Ensinar é ensinar a ler para que o aluno se torne capaz dessa apropriação, pois o conhecimento acumulado está escrito em livros, revistas, jornais, relatórios, arquivos. Ensinar é ensinar a escrever porque a reflexão sobre a produção de conhecimento se expressa por escrito.

De acordo com os autores, ensinar é ensinar a ler e a escrever. As duas capacidades estão interligadas, uma vez que uma é produto da outra e vice-versa. E ambas são responsabilidade da escola como um todo, não apenas da disciplina de Língua Portuguesa, especificamente. Dessa forma, é preciso que todas as disciplinas contribuam para o desenvolvimento dessas habilidades nos alunos, uma vez que o sucesso será mais rápido e abrangente, nesse caso.

Já com relação à escrita, os PCN trazem o texto como o centro das aulas de língua materna, visto que é por meio dele que todo e qualquer discurso se organiza. Segundo o documento:

O discurso, quando produzido, manifesta-se linguisticamente por meio de textos. O produto da atividade discursiva oral ou escrita que forma um todo significativo, qualquer que seja sua extensão, é o texto, uma seqüência verbal constituída por um conjunto de relações que se estabelecem a partir da coesão e da coerência. Em outras palavras, um texto só é um texto quando pode ser compreendido como unidade significativa global. Caso contrário, não passa de um amontoado aleatório de enunciados. (BRASIL, 1998, p. 21).

Dessa forma, os PCN tomam como objeto principal do ensino de Língua Portuguesa o texto. É por meio dele que se desenvolve todo o tipo de atividade, seja ela de leitura, de interpretação, de escrita, de produção ou de aspectos gramaticais. Em função disso, o documento ainda salienta que:

Tomando-se a linguagem como atividade discursiva, o texto como unidade de ensino e a noção de gramática como relativa ao conhecimento que o falante tem de sua linguagem, as atividades curriculares em Língua Portuguesa correspondem, principalmente, a atividades discursivas: uma prática constante de escuta de textos orais e leitura de textos escritos e de produção de textos orais e escritos, que devem permitir, por meio da análise e reflexão sobre os múltiplos aspectos envolvidos, a expansão e construção de instrumentos que permitam ao aluno, progressivamente, ampliar sua competência discursiva. (BRASIL, 1998, p. 27).

Assim, acreditamos que é principalmente no texto escrito que está a intersecção entre a leitura e a escrita. Esse texto, quando produzido pelo aluno, se manifesta, normalmente, na forma de redação escolar. Conforme Dargel (2011, p. 14):

Ao chegar à fase escolar, a criança estará apta, também, a lidar com textos escritos. Dependendo do tratamento que for dado ao léxico nas aulas de Língua Portuguesa, mais especificamente nas de leitura, essa criança, agora aluno, poderá ou não se tornar alguém com domínio de leitura e de produção de textos tanto orais como escritos.

A autora afirma que é na escola, principalmente, que a criança entra em contato regular com os textos escritos. A partir desse convívio é que ela vai desenvolver as habilidades de leitura e de escrita, objetivos principais do ensino de língua materna. Dargel (2011, p. 14) afirma que “para que esse objetivo seja

alcançado, há a necessidade de que os professores de Língua Portuguesa reservem, durante suas aulas, um espaço significativo destinado especificamente ao estudo do léxico”.

Dessa forma, consideramos que o ensino da língua como um todo engloba o ensino de vocabulário. Para que se consiga compreender um texto, é preciso entender as palavras que fazem parte dele, assim como para se escrever um texto, é preciso saber utilizá-las e combiná-las de forma coerente, relacionando-as e atribuindo sentido a elas. Então, concordamos com Rodrigues (2014, p. 69), que considera que:

[...] não podemos entender cabalmente um texto se não conseguirmos compreender as palavras desse texto, pois tal como defende Cruz (2007), a leitura não se restringe apenas à decodificação e reconhecimento das palavras, procurando fundamentalmente a compreensão da mensagem do texto.

Ou seja, acreditamos que ler é muito mais que decodificar letras, sílabas e palavras. Isso, conforme Guedes e Souza (2011, p. 01) é a penas o início do processo de leitura: “levar o aluno ao domínio do código escrito”. Depois que o aluno domina essa parte, é preciso que ele faça relações. Assim, compreenderá, de fato, o que está escrito no plano da palavra e no plano de sua articulação em um todo de sentido. Conforme Rodrigues (2014, p. 30):

Para que a atividade da leitura seja bem sucedida, não basta apenas ler as palavras, as frases de um texto, é também fundamental compreender o que está escrito. E para que se consiga compreender a mensagem de um texto escrito, é necessário não só reconhecer as palavras aí presentes, mas também a forma como se relacionam entre si.

Esse ideal vai ao encontro de nossa pesquisa, que acredita que a devida importância dada ao ensino de vocabulário, contextualizado, contribui para a formação de leitores e de escritores competentes. De acordo com Pais e Sardinha (2011, p. 01):

A incidência de crianças com dificuldades de aprendizagem, em geral, e com dificuldades ao nível da aprendizagem da leitura e da escrita, em particular, continua a ser uma realidade marcante do dia-a-dia das nossas escolas. Neste contexto, a investigação realizada nas últimas décadas veio demonstrar a importância do desenvolvimento técnico-didático das formas de abordagem aos processos de ensino explícito do vocabulário, enquanto

elemento e factor de aproximação das crianças aos livros e à leitura e a promoção do sucesso escolar.

Então, assim como os autores, acreditamos que é a partir de um vocabulário variado que o indivíduo se torna um leitor (e, conseqüentemente, um escritor) proficiente. A respeito do vocabulário, conforme Pais e Sardinha (2011, p. 02), “[...] quanto mais o sujeito leitor possuir, mais facilmente, ao ler, deduz do contexto o significado das palavras novas”. Ou seja, quanto maior for o repertório lexical de uma pessoa, com mais facilidade ela compreenderá novas palavras, por meio do contexto, deduzindo seu significado. Dessa forma, também compreenderá com maior facilidade o que está lendo, visto que as palavras desconhecidas são, muitas vezes, um entrave para a leitura. Nesse sentido, Rodrigues (2014, p. 71) argumenta:

[...] alguns estudiosos postulam que os leitores com dificuldades, ou os chamados “maus leitores”, desconhecem o significado da maioria dos vocábulos dos textos, sobretudo se se trata de palavras mais abstratas, complexas ou pouco habituais. Esta situação acontece muitas vezes com os alunos de zonas socioeconômicas desfavorecidas que dominam um conjunto muito reduzido de vocábulos, o que pode levar, a maioria das vezes, ao insucesso escolar.

Em contrapartida, o mesmo autor salienta que “se alguém apresenta um vocabulário vasto, multifacetado, em quantidade considerável e rico, enquadra-se no perfil do leitor hábil e esta capacidade terá uma grande influência na compreensão da leitura [...]” (RODRIGUES, 2014, p. 71). Ou seja, quanto menor o vocabulário de um indivíduo, mais dificuldades ele terá em ler e compreender. Por outro lado, quanto maior for seu repertório, mais facilidade ele terá nessa compreensão, obtendo sucesso em sua leitura. E esse sucesso é também o que irá abrir caminhos para que o indivíduo tome gosto pela leitura, tornando-se interessado, observador, curioso, questionador.

Além disso, de acordo com Pais e Sardinha (2011, p. 03), “[...] no âmbito da aquisição do vocabulário, se estabelecem nexos entre o que já se sabe e a informação nova”. Dessa forma, não só é possível deduzir o sentido de itens desconhecidos com mais facilidade, como também é possível expandir o seu vocabulário, por meio da aquisição desses novos itens, incorporados com êxito ao repertório do leitor, por estarem relacionados ao contexto em que estão inseridos.

Conforme Biderman (1998, p. 178-179), “por enorme que seja o léxico de uma língua, é reduzido o repertório efetivamente utilizado pelos falantes, até mesmo na

língua escrita, que é a variante da língua que se serve de um vocabulário mais rico e mais variado”. É nesse sentido que a incorporação de novas formas ao nosso vocabulário se dá por meio do uso, que só acontece quando de fato compreendemos seu sentido, sendo capazes de utilizar essas novas formas com propriedade. Com relação a isso, Dargel (2011, p. 14) coloca que:

Afirmamos isso por reconhecermos que o conhecimento da função discursiva dos itens lexicais se constitui como fator primordial na construção do sentido de um texto. Se o aluno for orientado pelo professor a como ampliar sua competência lexical haverá, indubitavelmente, acréscimo em seu repertório de palavras e também a ampliação de sua visão de mundo.

Sabemos que o vocabulário utilizado pelos alunos diz muito sobre a sua história e o contexto da comunidade em que estão inseridos, pois não é possível desvincular o léxico da cultura. Em função disso é que se deve aproximar os conteúdos e textos trabalhados em sala de aula com a realidade dos educandos, na medida em que, quanto mais próximos, mais significativos são os aprendizados.

Dargel (2011), ao trazer os textos como ponto central das aulas de língua materna, afirma ser necessário o trabalho com foco no vocabulário desses textos. O autor salienta o seguinte:

[...] constatamos que o texto se configura como o ponto central das aulas de Língua Portuguesa, mas que, em determinadas etapas da compreensão ou da produção do texto, faz-se necessário realizar um trabalho sistemático com a palavra. Contudo, não de forma descontextualizada. Estudar uma palavra é ver todas as suas possibilidades de uso e significações, de forma e de seus contrastes. Enfim, algo bem mais amplo que saber apenas o sinônimo dessa palavra e a sua ortografia. (DARGEL, 2011, p. 15).

Nesse sentido está inserido o foco de nosso estudo: o ensino de vocabulário como forma de expandir as capacidades de leitura, interpretação e compreensão (e, conseqüentemente, escrita) dos alunos. É a partir da expansão do seu vocabulário ativo, ou seja, daquele repertório que os educandos de fato utilizam, que eles desenvolverão também suas habilidades interpretativas. Por isso, concordamos com a autora, que afirma que:

[...] a ideia de que se o ensino do vocabulário for realizado de maneira eficiente, haverá a ampliação do conhecimento lexical do aluno e, conseqüentemente, uma melhora na qualidade da aprendizagem da Língua Portuguesa como um todo. A pessoa detentora de um melhor repertório lexical manifesta maior poder do mundo referencial e, assim, não apenas

compreende de forma mais eficaz o que lhe é dito [...]. (DARGEL, 2011, p. 77).

Ao encontro do que trata Dargel está o pensamento de Rodrigues (2014, p. 69), que afirma:

existe uma relação estreita entre o vocabulário e a leitura que pode ser entendida de uma forma circular: o conhecimento lexical é importante para a leitura e a prática da leitura é uma forma de se conseguir desenvolver esse conhecimento lexical.

Então, a ligação entre leitura e repertório vocabular é cíclica, uma vez que um participa e contribui para o sucesso do outro. Assim como a leitura nos permite ampliar nossos horizontes e, conseqüentemente, nosso vocabulário, por sua vez, o nosso repertório vocabular também nos permite ler com proficiência.

Além disso, segundo o mesmo autor, “[...] já ficou documentado que o domínio do vocabulário potencia o desempenho da leitura, a leitura constitui um meio que possibilita a aprendizagem de palavras novas, o que significa que quanto mais se lê, maior é o desenvolvimento lexical” (RODRIGUES, 2014, p. 72). Nesse sentido está a importância da formação de leitores assíduos e praticantes, pois quanto mais se lê, mais se entra em contato com palavras novas e com diferentes significados de palavras já conhecidas. É dessa forma que expandimos nosso repertório vocabular e é por isso que é tão importante a maneira como o aprendiz entra em contato com a leitura, conforme já mencionado no início deste capítulo, já que é normalmente nesse início da relação que se estabelece o sentimento do leitor com a leitura.

Por fim, Rodrigues (2014, p. 70) afirma que “o papel do vocabulário é muito relevante no mundo da leitura e as palavras estão intimamente ligadas à ação de ler”. É com esse pensamento que, após nossa pesquisa, ao final deste estudo, apresentamos propostas de atividades de ensino de vocabulário, com ideias que podem ser aplicadas em sala de aula de Ensino Fundamental, objetivando justamente o alargamento do repertório vocabular que os alunos já trazem consigo. Complementando esse pensamento:

É também consensual a ideia de que existe uma forte influência do vocabulário na compreensão de leitura. O indivíduo que tem bons conhecimentos ao nível da língua e do conhecimento lexical tem fortes probabilidades de apresentar um bom desempenho na compreensão de

leitura (GIASSON, 1993) e no seu desenvolvimento da linguagem. (RODRIGUES, 2014, p. 70).

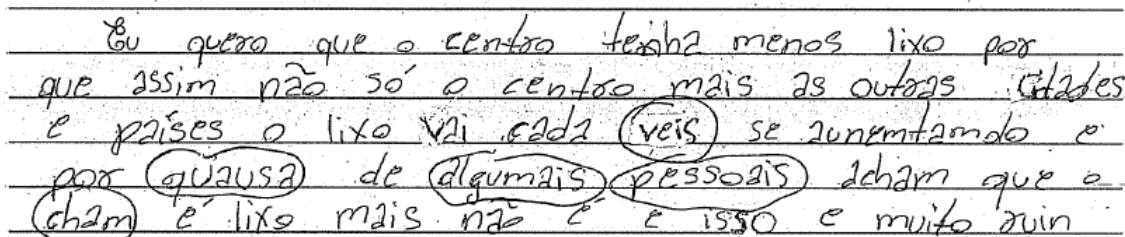
É dessa forma e com esse pensamento que seguimos nosso trabalho. O próximo capítulo apresenta o experimento realizado. Inicialmente, explicamos detalhadamente os métodos que foram empregados e o nosso *corpus* de estudo: as redações dos alunos do Ensino Fundamental de escolas públicas. Após, mostramos os estudos piloto, desenvolvidos com amostras de nosso *corpus*, por meio de comparações entre as duas redes de ensino e com os *corpora* de contraste. Ao final do capítulo, trazemos o experimento principal, com todo *corpus* de 154 redações escolares, contrastando-as com os *corpora* de referência.

## 5 EXPERIMENTO DESCRITIVO E ANALÍTICO

Com vistas a descrever o vocabulário verificado na produção escrita dos estudantes das duas escolas, o *corpus* principal de estudo deste trabalho é composto por uma coleção de 154 redações manuscritas. Conforme já citado, esse *corpus* foi constituído pela professora de Língua Portuguesa dos estudantes ao longo de oito anos de docência nessas escolas e reunido como um acervo particular. O período de produção dos textos do acervo selecionados especificamente para este estudo corresponde aos anos de 2017, 2018 e 2019. Os textos foram armazenados pela professora em formato de fotocópias. Parte desse acervo pessoal tornou-se nosso material de estudo.

Para tanto, os textos foram desidentificados e numericamente codificados conforme a escola de origem, a temática do texto e o ano/ciclo ao qual pertenciam os redatores. Em seguida, os manuscritos foram digitados em editor de texto no formato *.doc*, conforme a Figura 3 a seguir busca ilustrar.

Figura 3 – Trecho de texto original manuscrito II



Eu quero que o centro tenha menos lixo por  
que assim não só o centro mais as outras cidades  
e países o lixo vai cada (veis) se aumntando e  
por (quausa) de (algumais) (pessoais) acham que o  
(cham) é lixo mais não é e isso e muito ruin

Fonte: Acervo pessoal.

Quadro 5 – Trecho de texto original digitado II

Eu quero que o centro tenha menos lixo por que assim não só o centro mais as outras cidades e países o lixo vai cada veis se aumntando e por quausa de algumais pessoais acham que o cham é lixo mais não é e isso e muito ruin
---

Fonte: Acervo pessoal.

A Figura 3 é o original, manuscrito, de trecho de um dos textos dos alunos. As marcas em torno de algumas palavras são de autoria da professora. Logo abaixo, no Quadro 5, seu correspondente em formato *.doc*. Conforme visualizamos, os eventuais erros de ortografia, acentuação, pontuação, concordância, entre outros, foram mantidos/preservados, para que fosse possível o mapeamento do vocabulário



escrito desses alunos de acordo com suas características originais. Exemplo de uma redação completa de cada escola pesquisada encontra-se no Anexo I deste trabalho.

## 5.1 MATERIAIS

Considerando o conjunto total de textos da amostra fixada, reunidos em arquivo único, temos, conforme Beber Sardinha (2004), um *corpus* de estudo de tamanho pequeno, com 26.424 *tokens* (total de palavras) e 4.113 *types* (palavras diferentes). Ainda assim, conforme já explicado no Capítulo 2, o *corpus* é representativo do perfil de vocabulário escrito desses estudantes de Ensino Fundamental, uma vez que é composto por redações criteriosamente selecionadas a partir de uma população de textos que corresponde a redações armazenadas ao longo dos oito anos de docência em escolas públicas.

Diante do acervo à disposição, com temáticas e quantidades variadas de textos, optamos por selecionar redações de maneira que a amostra ficasse homogênea em relação às duas realidades pesquisadas. O objetivo era de que a heterogeneidade de temas tratados nas redações fosse um fator a ser controlado frente ao objetivo de se desenhar um perfil de vocabulário.

Dessa forma, nossa preferência foi por selecionar os textos da mesma temática produzidos nas duas escolas. Entre as cinco propostas disponíveis, duas eram iguais. A partir dessas duas propostas, os dois primeiros estudos piloto descritos a seguir foram desenvolvidos.

Outra maneira de balancear nosso *corpus* foi selecionar textos com número de linhas e com avaliações semelhantes. Sobre isso, vale destacar que “[...] a sistemática de avaliação é diferente em cada uma das escolas. Na escola estadual, **a avaliação que os textos receberam pela professora** segue uma escala numérica, enquanto na escola municipal é utilizado um sistema de conceitos descritivos” (SILVA, 2020, p. 34, grifo nosso). Ainda assim, optamos por aqueles que representavam as diversas notas possíveis – para que a amostra ficasse variada – e que equivalassem, em termos de avaliação, aos dois modelos. Além disso, a quantidade de redações que compõe os grupos por escola e por etapa escolar foi a mesma.

Cabe ressaltar também as condições de produção desses textos. Igualmente, nas duas escolas, foram trabalhados textos de apoio e questões de interpretação em todas as propostas. Esses textos foram crônicas (ANEXO III), todas coincidentemente publicadas em jornais, ou filmes (ANEXO V). Após as discussões e interpretações realizadas em sala de aula, os alunos produziram suas redações de acordo com as orientações (ANEXO II). Eles tiveram cerca de 90 minutos, o equivalente a dois períodos, para produzirem seus textos, que foram entregues escritos à mão, presencialmente, na aula de Língua Portuguesa. Os textos foram avaliados, conforme o sistema avaliativo de cada instituição, fotocopiados sem identificação nominal e devolvidos aos respectivos redatores.

As 154 redações selecionadas do acervo pessoal foram, então, processados em arquivo único, com o apoio do *software AntConc 3.5.8* (ANTHONY, 2019). O Quadro 6, a seguir, resume algumas características gerais desse *corpus* no tocante à distribuição e frequência de itens. Os elementos assinalados com \* (asterisco) correspondem a contagens a partir de listas lematizadas. Nesse processo, realizado por uma ferramenta computacional, as diferentes flexões de uma mesma forma-base são reunidas sob único item. O modo de geração das formas lematizadas do *corpus* está explicado mais adiante. Foram desprezadas, para as contagens geradas, as diferenças maiúscula/minúscula:

Quadro 6 – Composição e características do *corpus* de estudo – amostra sob exame

<b>Composição total</b>	154 redações (05 temas / textos das duas escolas)
<b>Composição por escola</b>	77 redações de cada escola
<b>Tokens (Total de palavras)</b>	26.424
<b>Types (Palavras diferentes)</b>	4.113
<b>Riqueza Lexical (Type/Token Ratio)</b>	13,22%
<b>Palavras gramaticais mais frequentes (amostra)</b>	o, que, e, a, de, eu, para, um, as, os, mais, com, no, se, uma, por, em, do, na, eles, minha, da, muito, ele, mas, sem, meu, ou, todos, quando, depois, isso, bom, porque, tudo, me, como
<b>Palavras lexicais mais frequentes (amostra)</b>	Brasil, pessoas, quero, filme, tem, casa, dia, ter, ser, vida, coisas, futuro, seria
<b>*Hapax (palavras de ocorrência única): Letra A (amostra)</b>	abaixar, abandonar, abc, aberto, abolir, abortar, abrigo, abuso, acaso, acender, acima, acontecido, acontecimento, acordado, acorrentado, acostumar, acre, acumulado, acumular, adaptar
<b>*Hapax (palavras de ocorrência única): Letra P (amostra)</b>	pago, pagamento, paisagem, palavrar, palha, palmar, papa, papo, par, parcelar, particularmente, parto, passe, passeio, passo, pato, peculiaridade, pedofilia, pedrinho, pedro

Fonte: Elaboração própria.

Ao observar a listagem gerada, as palavras mais frequentes e os *hapax*, percebemos que eventuais erros de grafia foram preservados, como “palavrar”, por exemplo. Além disso, o Quadro 6 também nos mostra algumas palavras de mesma base que não foram agrupadas pelo lematizador, como: “acumular-acumulado”, “acontecido-acontecimento”, “pedro-pedrinho”. Sabemos que um processo de lematização automática pode apresentar falhas, mas, ainda assim, optamos por utilizá-lo, sobretudo porque a ferramenta se mostrou satisfatória para nosso estudo, apresentando número de acertos infinitamente superior à quantidade de falhas observadas.

Considerando dois grupos de redações, um por escola, há apenas 2 temas em comum entre as 5 propostas de produção escrita que compõem nossa pesquisa (ANEXO II). Um desses temas coincide nos dois sextos anos das duas escolas e o outro nos dois oitavos anos.

Além dessas características, o *corpus* amostral reunido pode ser classificado, segundo Berber Sardinha (2004), conforme o Quadro 7:

Quadro 7 – Classificação do *corpus*

<b>Modo</b>	Escrito (redações escolares)
<b>Tempo</b>	Sincrônico / Contemporâneo
<b>Seleção</b>	De amostragem / Estático/ Equilibrado
<b>Conteúdo</b>	Regional (redações de alunos de escolas públicas do RS)
<b>Autoria</b>	Mista (alunos imigrantes haitianos e venezuelanos)
<b>Finalidade</b>	De estudo (análise contrastiva do vocabulário escrito de estudantes de escolas municipal – POA – e estadual – RS)
<b>Fechamento</b>	Fechado

Fonte: Adaptado de Grama (2016, p. 119).

Nosso *corpus* de estudo, ainda conforme Berber Sardinha (2004), possui função representativa, ou seja: “na sua essência, um *corpus*, seja de que tipo for, é tido como representativo da linguagem, de um idioma, ou de uma variedade dele” (BERBER SARDINHA, 2004, p. 22). Dessa forma, conforme já exposto no Capítulo 2, acreditamos na validade da amostra que compõe nosso *corpus*, uma vez que, enquanto docente e de acordo com nossa experiência nessas duas realidades de ensino público, sabemos que a amostra é representativa desses alunos.

Além de representativo, o *corpus* de estudo pode ser categorizado de diferentes maneiras: por temas, por ano/ciclo e por escola/instituição de origem dos estudantes produtores dos textos. Isso é o que sintetizamos a seguir:

Quadro 8 – Temáticas do *corpus* de estudo

<b>POR TEMAS: cinco temas diferentes</b>				
TEMA 1: A vida como filme  36 redações	TEMA 2: Um sonho possível 15 redações	TEMA 3: Voltar de férias é... 14 redações	TEMA 4: Os Goonies 14 redações	TEMA 5: O Brasil que eu quero... 75 redações

Obs.: Todas as propostas dos cinco temas estão reproduzidas na íntegra em Anexo II.

Fonte: Elaboração própria.

Quadro 9 – Ano/ciclo do *corpus* de estudo

<b>POR ANO/CICLO: dois anos/ciclos diferentes</b>	
6º ANO/B30: 60 redações	8º ANO/C20: 94 redações

Fonte: Elaboração própria.

Quadro 10 – Escolas do *corpus* de estudo

<b>POR ESCOLA: duas escolas diferentes</b>	
ESCOLA A (municipal): 77 redações	ESCOLA B (estadual): 77 redações

Fonte: Elaboração própria.

Além do *corpus* amostral em estudo, como contraste, reunimos os textos que serviram de apoio para os estudantes em meio às propostas de produção escrita (quando existentes). Também integram o material de contraste os textos dos livros didáticos de Língua Portuguesa lidos ou estudados pelos redatores antes da sua produção ser feita. Todos esses materiais são chamados, nesta pesquisa, de *corpora* de contraste. No cenário da Linguística de Corpus, naturalmente, trabalha-se com *corpus* de estudo e *corpus* de referência. No entanto, recorreremos a um material de apoio como esse para que fosse possível observar eventuais influências sobre o vocabulário manifestado nas redações escolares. Uma amostra desses *corpora* de contraste encontra-se na seção de Anexos.

Os materiais de contraste são compostos, primeiramente, por três textos de apoio (ANEXO III), lidos, discutidos e interpretados pelos alunos imediatamente antes das produções textuais. Em função da proximidade da leitura desses textos com a produção escrita dos alunos, entendemos que o vocabulário desses textos poderia ser associado à memória de curto prazo dos estudantes. Apenas três das cinco propostas de produção apresentaram texto de apoio em função de que as outras duas eram propostas relacionadas à interpretação de filmes, então, os

educandos assistiram ao filme antes da escrita (em vez de lerem o texto de apoio). Abaixo estão as características de cada um desses três textos de apoio:

Quadro 11 – Textos de apoio para os redatores – usados como contraste

	<b>Texto de Apoio 1</b>	<b>Texto de Apoio 2</b>	<b>Texto de Apoio 3</b>
<b>Título</b>	A vida como filme	Voltar de férias é...retomar a rotina	Futuro
<b>Escola</b>	ambas	Municipal	Estadual
<b>Autor</b>	Moacyr Scliar	Nathália Prestes Bosak	Nilson Souza
<b>Ano</b>	2003	2012	2018

Fonte: Elaboração própria.

Na seção de Anexos deste trabalho também estão esses textos de apoio na íntegra (ANEXO III).

Os filmes referentes às outras duas propostas que não tiveram texto de apoio são: “Os Goonies” e “Um sonho possível”. Os resumos desses filmes também se encontram nos Anexos (Anexo V).

Conforme já mencionado, também foram utilizados, para contraste do vocabulário, textos dos livros didáticos de Língua Portuguesa das duas escolas (ANEXO IV). Esses textos foram considerados por terem sido trabalhados em sala de aula com os alunos antes das produções escritas. Vale salientar que isso ocorreu antes dos redatores terem tido contato com os textos de apoio para a produção escrita solicitada. Em função dessa questão temporal, entendemos que o vocabulário dos textos dos livros didáticos poderia estar sendo aproveitado nas redações, associado à memória de longo prazo dos estudantes.

Os textos desses livros didáticos também se encontram nos Anexos deste trabalho e foram os seguintes:

Quadro 12 – Textos didáticos utilizados como contraste

	<b>Texto didático 1</b>	<b>Texto didático 2</b>
<b>Título</b>	A cara da vida moderna	Um bom sujeito (fragmento)
<b>Escola</b>	Municipal	Estadual
<b>Autor</b>	Walcyr Carrasco	Antônio Carlos Olivieri
<b>Livro didático</b>	Português – Linguagens	Tecendo Linguagens
<b>Ano</b>	2014	2015

Fonte: Elaboração própria.

Quanto a esses materiais dos livros didáticos previamente utilizados pelos redatores, importa frisar que a sua utilização, para verificação de alguma influência

sobre o léxico das redações, foi reavaliada e descartada. Isso será explicado na seção de estudos piloto deste capítulo.

Como *corpus* de referência, isto é, como um *corpus* geral que representa a língua escrita do Brasil e que não inclui o *corpus* de estudo, utilizamos, no experimento principal, o CorPop<sup>19</sup> (PASQUALINI, 2018). Esse é um conjunto de textos que serve como referência do português popular escrito. Foi selecionado por ser atual, por espelhar um uso do português popular escrito e por ter sido produzido em meio às atividades dos nossos grupos de pesquisas.<sup>20</sup> Os textos que compõem esse *corpus* (PASQUALINI, 2018) são:

Figura 4 – Dados do CorPop

Módulo	Types Palavras diferentes	Tokens Total de palavras - formas
PorPopular Diário Gaúcho	6.378	30.944
Jornal Hora de Santa Catarina	4.118	18.303
Jornal Boca de Rua	8.913	71.454
Jornal Diário da Causa Operária	7.841	59.785
Textos de estudo Retratos da Leitura no Brasil	22.463	430.806
Textos da Coleção "É Só o Começo"	8.161	73.507
Total	32.138	684.799

Fonte: Paqualini (2018).

Nosso *corpus* de referência, CorPop, reúne textos de acordo com o nível de letramento médio dos leitores do país verificado entre 2015 e 2016. Esse nível de letramento está de acordo com dados recolhidos nas pesquisas Indicador de Analfabetismo Funcional (INAF)<sup>21</sup> e Retratos da Leitura no Brasil,<sup>22</sup> e com questionários realizados com leitores.

<sup>19</sup> Dados desse *corpus*, como a lista de palavras, estão disponíveis em: <<http://www.ufrgs.br/textecc/porlexbras/corpop/index.php>>. Acesso em: 20 fev. 2020.

<sup>20</sup> Dados e recursos produzidos pelos grupos de pesquisa podem ser encontrados em: <<http://www.ufrgs.br/textecc/acessibilidade/>> e <<http://www.ufrgs.br/textecc/porlexbras/>>.

<sup>21</sup> INAF: Indicador de Analfabetismo Funcional. É uma pesquisa que tem o objetivo de mensurar o índice de analfabetismo da população brasileira entre 15 e 64 anos. Essa pesquisa avalia habilidades e práticas de leitura, de escrita e de matemática. Dados retirados do site: <<https://ipm.org.br/inaf/>>. Acesso em: 20 fev. 2020.

Num segundo momento, ao analisarmos as *keywords* e medidas de chavidade, utilizamos outros *corpora* de referência: o Banco de Português, o Lácio Web e o Corpus Brasileiro. Os dois últimos *corpora* estão disponíveis na internet.

O Banco de Português<sup>23</sup> é um *corpus* de linguagem geral do português, fruto “[...] do projeto DIRECT (LAEL, PUCSP), com mais de 230 milhões de palavras, de fontes orais e escritas [...]” (BERBER SARDINHA, BÁRBARA, 2005, p. 97). Hoje, somente está acessível incorporado ao Corpus Brasileiro, disponível na Linguateca (<https://www.linguateca.pt/ACDC/>). Em nosso estudo, foram usados os 100 mil primeiros *tokens* do Banco de Português como referência para a identificação de palavras-chave.

O Lácio Web<sup>24</sup> é um projeto que divulga e disponibiliza na *web* vários *corpora* do português brasileiro escrito contemporâneo, além de ferramentas linguístico-computacionais. É subdividido em *subcorpora* que representam diversos gêneros e tipos textuais. Entre eles, usamos o Lácio-Ref, *corpus* aberto e de referência do Lácio Web, composto por textos (informativos, literários, científicos, jurídicos e institucionais) em português, respeitando a norma culta. Do Lácio-Ref, utilizamos uma lista com 7.270.715 *tokens* para identificação de palavras-chave no nosso *corpus* de redações.

O Corpus Brasileiro<sup>25</sup> (BERBER SARDINHA; MOREIRA FILHO; ALAMBERT, 2010) também é um *corpus* contemporâneo e de língua geral, que visa disponibilizar on-line uma coletânea de aproximadamente um bilhão de palavras do português brasileiro contemporâneo, composto por diferentes tipos de linguagens. Esse *corpus* é resultado de projeto coordenado por Tony Berber Sardinha, que ainda está em desenvolvimento pelo Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem (LAEL) da PUC-SP. Esse *corpus* tem 92,4% do componente escrito e 7,6% do componente falado. Mais informações sobre o Corpus Brasileiro

---

<sup>22</sup> Retratos da Leitura no Brasil: pesquisa que avalia os impactos e orienta políticas públicas do livro e da leitura. Seu objetivo é melhorar os indicadores de leitura do brasileiro. Dados retirados do site: <[http://prolivro.org.br/home/images/2016/Pesquisa\\_Retratos\\_da\\_Leitura\\_no\\_Brasil\\_-\\_2015.pdf](http://prolivro.org.br/home/images/2016/Pesquisa_Retratos_da_Leitura_no_Brasil_-_2015.pdf)>. Acesso em: 20 fev. 2020.

<sup>23</sup> Disponível em: <<https://www.pucsp.br/pesquisa-seleta-2011/projetos/047.php>>. Acesso em: 08 jul. 2020.

<sup>24</sup> Disponível em: <<http://143.107.183.175:22180/lacioweb/index.htm>>. Acesso em: 08 jul. 2020.

<sup>25</sup> Disponível em: <<http://corpusbrasileiro.pucsp.br/cb/Acesso.html>>. Acesso em: 08 jul. 2020.

podem ser consultadas no site da Linguateca.<sup>26</sup> Em nosso estudo, foram usados 3.674.746 *tokens* desse *corpus* para identificação de palavras-chave.

## 5.2 ETAPAS DO EXPERIMENTO DESCRITIVO

Para que pudesse ser examinado, o acervo de redações manuscritas foi digitado em um editor de texto. Esse trabalho de digitação foi minucioso e exaustivo. Além disso, foi preciso não só desabilitar o corretor automático do computador, como também cuidar algo que parecia ser inevitável: realizar correções. Após a digitação, foi feita uma conferência geral do material produzido de modo a se preservar a apresentação original do texto conforme escrito pelo(a) aluno(a).

Depois disso, foi necessário organizar e categorizar as redações de acordo com a escola e a temática do texto, nomeando os arquivos um a um, de maneira que fosse possível identificar essas duas informações. Assim, o *corpus* de estudo, após armazenado em arquivo *.doc*, foi convertido para o formato *.txt*, sem formatação, para que fosse possível processá-lo em ferramentas computacionais – tais como os geradores de listas e de frequências de palavras.

O *corpus* organizado foi processado com o programa *AntConc* 3.5.8 (ANTHONY, 2019), um *software* de acesso livre, que contém ferramentas que nos permitem gerar dados estatísticos (como listas de palavras e de contextos) a partir de um texto em formato digital. Essa ferramenta originou dados como o número de *tokens* e de *types*, as frequências das palavras e o *hapax legomena*. A partir desses dados, contabilizamos: a) a riqueza lexical – relação *types/tokens*; e b) proporção de palavras lexicais e gramaticais presentes nos textos.

Também os materiais do *corpus* de contraste, tanto os textos de apoio, quanto os textos dos livros didáticos, foram submetidos ao *AntConc*. A partir dos dados obtidos por meio da ferramenta, fizemos pequenos estudos piloto, com segmentos menores do nosso *corpus* total de estudo. O número de textos pesquisados foi aumentando a cada novo estudo piloto, a fim de observarmos as mudanças nos resultados à medida que diversificávamos a quantidade e a temática das redações. Isso também foi um recurso para uma melhor delimitação do *corpus* amostral que estaria sob exame.

---

<sup>26</sup> Disponível em: <[https://www.linguateca.pt/Repositorio/manual\\_cb.pdf](https://www.linguateca.pt/Repositorio/manual_cb.pdf)>. Acesso em 08 jul. 2020.

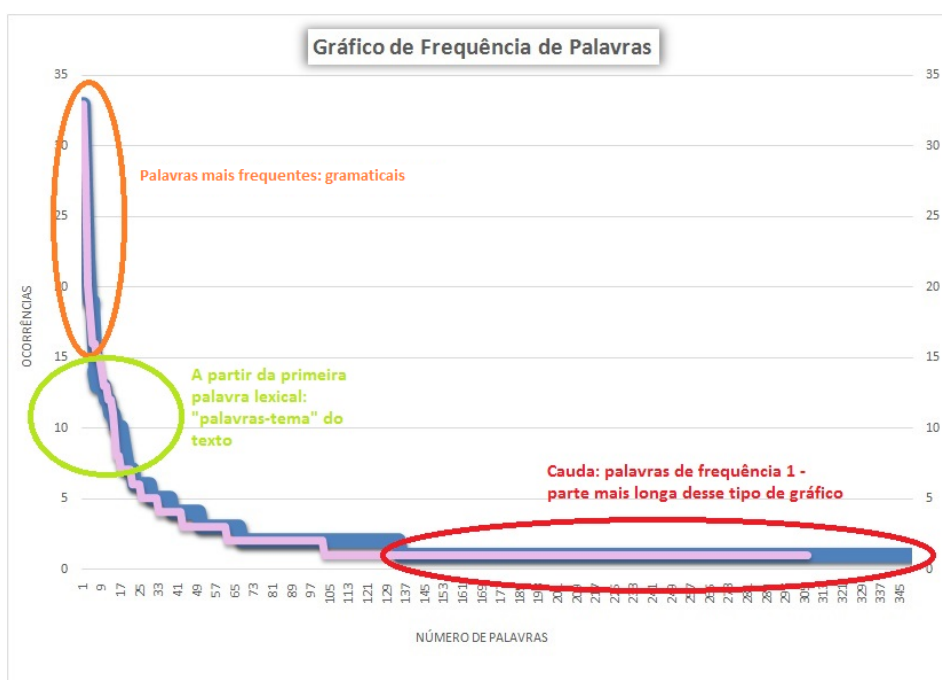


Nesses estudos piloto, os textos do *corpus* que foram usados em cada um e os resultados são detalhados nas seções a eles dedicadas. Após os estudos com pequenos grupos de textos, realizamos o experimento principal, fixando-se 154 textos como *corpus*. O experimento principal está descrito na seção 5.4.

A fim de investigar a validade de nossas hipóteses, observamos, em todos os estudos piloto e no experimento principal, o número total de palavras (*tokens*), o número de palavras diferentes repetidas (*types*), os graus de frequência das palavras, a proporção de emprego de palavras lexicais e gramaticais, a riqueza lexical (*type/token ratio*) e as palavras de frequência única (*hapax legomena*). Todos esses dados, das duas escolas, foram comparados entre si e com os *corpora* de contraste nos estudos piloto. Após os pequenos estudos, no experimento principal, com todos os textos da amostra reunidos, o diferencial foi o uso do CorPop como *corpus* de referência.

Uma etapa importante para a descrição de frequências e distribuições do vocabulário por tipos de palavras foi a geração de diferentes gráficos de distribuição lexical. Esses gráficos, gerados no programa *Microsoft Office Excel*, reproduzem o que ocorre nas listagens, tanto a quantidade total de palavras, quanto o que chamamos de camadas, e são compostos de acordo com a seguinte sistemática:

Gráfico 2 – Camadas de um gráfico de frequência de palavras



Fonte: Elaboração própria.

Conforme vemos no Gráfico 2, as palavras mais frequentes são itens gramaticais. Essas unidades formam uma a primeira camada (em laranja) de frequência. Nessa camada, temos, em geral, dados sobre a gramática da língua envolvida, em função de haver apenas um repertório de palavras gramaticais.

Em seguida, a partir da primeira palavra lexical (em verde), estão as palavras-tópico do texto. Essas palavras aparecem também com bastante frequência em função de serem os temas dos textos analisados. E, por fim, está o que chamados de “cauda” (em vermelho), camada geralmente mais extensa. Nessa camada, temos o *hapax*, com as palavras de frequência única, ou seja, que ocorrem apenas uma vez no *corpus*. Essas palavras aparecem apenas uma vez justamente porque demonstram as especificidades de um texto ou mesmo de determinado autor, dependendo do *corpus* analisado. Sobre isso tratamos ainda, mais adiante.

À medida que fomos progredindo em nossos estudos piloto, constatamos que a camada de itens de frequência única (*hapax legomena*) era a mais promissora justamente por espelhar as especificidades de cada grupo. A partir dos itens de frequência única, imaginamos que seria possível descobrir semelhanças ou diferenças nas escolhas particulares do vocabulário dos alunos.

Além disso, seria possível avaliar a incidência dos *hapax* que correspondiam a grafias de palavras com erros. Nos estudos piloto, percebemos que uma mesma palavra-base, escrita de duas ou três formas diferentes (com s, ss ou ç, por exemplo), contabilizava palavras diferentes e, muitas vezes, palavras de única frequência.

Fora isso, os verbos e suas várias conjugações possíveis, e os substantivos e adjetivos e suas flexões de gênero e de número, também só apareciam no conjunto dos itens *hapax* por não estarem lematizados. Em função disso, percebemos que nosso tratamento lexical-estatístico precisava ser revisto. Nesse sentido, foi necessário parar e repensar, pois tais questões – a variação ortográfica por erros e as flexões morfológicas – estavam afetando nossos resultados.

Foi então que sentimos a necessidade de lematizar as listas de palavras do *corpus* de estudo, considerando, a partir de então, apenas a forma básica ou canônica das palavras. Conforme Cúrcio (2006, p. 11):

A lematização é um processo que atua como uma espécie de filtro, deixando o número total de formas de uma maneira que evite a sua repetição. Como, por exemplo, as diferentes conjugações de um mesmo verbo, ou as flexões que indicam número e gênero, etc. A lematização é uma maneira de garimpar o texto, de modo que as formas sejam contabilizadas mais estritamente.

Para esclarecer o processo, consideramos a seguinte frase: “Seus filhos serão como foram seu pai”. Essa frase, que contém 7 diferentes palavras/*tokens*, numa lista lematizada de itens ficaria com 5 diferentes itens, de diferentes frequências: seu(2), filho(1), ser(2), como(1), pai(1). Isso acontece porque o pronome “seu” – que aparece flexionado no plural – e o verbo “ser” – que aparece conjugado no futuro e no passado – entram na contagem de suas formas básicas: *seu*, *ser*, com duas ocorrências cada um. Dessa forma, muitas das palavras que até então compunham nosso *hapax*, aparecendo uma única vez, ao serem lematizadas, foram atribuídas a sua forma canônica, não aparecendo mais junto com as palavras de única ocorrência.

Em função disso, quando tivemos a parceria de colegas do NILC-USP,<sup>27</sup> Prof. Dr. Thiago Pardo e seu orientando de Iniciação Científica, Angelo Tessaro, lematizamos nosso *corpus* de estudo, algo que aconteceu a partir do terceiro estudo piloto. Esses cientistas da computação nos auxiliaram por meio da criação de um programa que lematizou as nossas listas de palavras. Esse programa parte dos dicionários embutidos no programa UNITEX,<sup>28</sup> com suas respectivas formas canônicas associadas. Esse léxico pode ser encontrado no seguinte link: <<http://www.nilc.icmc.usp.br/nilc/projects/unitexpb/web/dicionarios.html>> e funciona da seguinte maneira:

1) Foi criado um *script* para alterar o léxico e manter no arquivo apenas a forma flexionada e a forma canônica, sem os detalhes da flexão, que iriam interferir na contagem.

---

<sup>27</sup> Centro Interinstitucional de Linguística Computacional da Universidade de São Paulo, USP, que fica em São Carlos. Seus projetos incluem estudos sobre o processamento computacional de linguagens naturais e a construção de recursos, ferramentas e aplicações, além de parceria com a indústria para transferência de conhecimento e tecnologia. Dados retirados do site: <<http://www.nilc.icmc.usp.br/nilc/index.php>>. Acesso em: 08 abr. 2020.

<sup>28</sup> O UNITEX/GramLab é uma suíte livre, multiplataforma, plurilíngue, baseada em léxicos e gramáticas para o processamento de corpora textuais. Dados retirados do site: <<https://unitexgramlab.org/pt>>. Acesso em: 08 abr. 2020.

2) Foi criado um *script* que procurava cada uma das palavras flexionadas da lista de palavras do léxico gerado no passo 1, guardava a contagem original, mas agora associada à forma canônica. Cada vez que a forma canônica se repetisse, era somado o valor da contagem da palavra flexionada que estava sendo procurada. Sendo assim, no fim, obtivemos um arquivo com as formas canônicas associadas a uma contagem, que é a soma de todos que possuem a mesma forma canônica no arquivo original.

3) Já que o arquivo gerado no passo 2 não seguiu um padrão ordenado, foi criado também um *script* que ordenasse as formas canônicas das mais frequentes para as menos frequentes, até que chegasse nas de frequência 1, a partir da qual a ordenação se tornou alfabética.

A lista que nos foi remetida corresponde ao procedimento do item 3, com ordenação por frequência, até as de frequência um, quando seguiu a ordem alfabética. A Tabela 1 abaixo exemplifica essa lista de palavras lematizada, com as 20 primeiras palavras mais frequentes no *corpus* de estudo:

Tabela 1 – Início da lista de itens lematizados do *corpus* de estudo

Palavra	Ocorrências	Palavra	Ocorrências
1. o	3089	11. para	426
2. que	1211	12. mais	349
3. e	1175	13.com	294
4. eu	962	14. ir	286
5. um	657	15. ser	259
6. de	622	16.querer	238
7. ele	534	17. brasil	227
8. no	519	18.pessoa	224
9. do	481	19. se	217
10. ter	466	20. muito	217

Fonte: Elaboração própria.

Naturalmente, esse processo de lematização, em sendo automático, comporta alguns erros. Todavia, acreditamos que os acertos superam esses erros e, por isso, utilizamos esses dados do *corpus* lematizado quando observados os itens que compõem o grupo *hapax* a partir do terceiro estudo piloto.

Com essa lematização pronta, foi possível seguir com nossos estudos, de maneira que os resultados ficassem mais precisos. Assim, seria possível contrastar de fato os dados e, sobretudo, a camada dos itens de frequência única.

À medida que avançávamos em nosso estudo, percebemos também a necessidade de analisarmos as *keywords*, ou seja, as palavras-chave de nosso *corpus* de estudo em relação a um *corpus* de referência. Conforme já mencionado nos materiais, os *corpora* usados como referência para o uso dessa ferramenta foram: Banco de Português, Lácio Web e Corpus Brasileiro.

Esse contraste foi feito por meio da ferramenta KeyWords, do programa *WordSmith Tools*. Conforme já mencionado no Capítulo 2, o *software* permite o uso de diferentes ferramentas para análise de *corpus*. Comparamos a lista de frequências de palavras do nosso *corpus* de estudo com diferentes *corpora* de referência. O resultado obtido mostra quais palavras empregadas nas redações são específicas em relação ao todo da língua portuguesa escrita culta. Nosso estudo com *keywords* foi realizado ao final do experimento principal, estando explicado na seção 5.5 deste capítulo.

A seguir, relatamos os resultados desses experimentos parciais. Esse relato se justifica pelo fato de que esses estudos parciais nos permitiram melhor fixar a amostra final sob exame. Depois deles está o nosso experimento principal, com todos os textos reunidos, e, ao final, a síntese dos resultados obtidos.

### 5.3 PARTE I: ESTUDOS PILOTO

Conforme já mencionado, o *corpus* de estudo foi categorizado por escola, por temática e por ano/ciclo. Então, a partir dessas categorizações, realizamos uma série de pequenos ensaios descritivos. Isso foi feito conforme indicações constantes no trabalho de Biber (2012, p. 43), que nos ensina que:

Outras questões relacionadas ao planejamento, no entanto, dependem de um corpus-piloto de textos para as investigações preliminares. [...] Independentemente do planejamento inicial, a compilação de um corpus representativo deve ser executada de maneira cíclica: um corpus-piloto deve ser compilado primeiro, representando uma gama relativamente ampla de variação, mas também representando um aprofundamento em alguns registros e textos.

Dessa forma, o primeiro estudo piloto foi realizado com apenas quatro textos por escola, que representavam 20% do total de textos reunidos sobre uma mesma temática. O segundo analisou todos os dezoito textos da mesma temática do

primeiro estudo, uma das poucas que se repete nas duas escolas. Já o terceiro estudo reuniu todos os textos do 8º ano/C20 disponíveis em nosso *corpus*, num total de 47 textos por escola, com temáticas variadas e, na maioria, não coincidentes entre os dois grupos. Por fim, o quarto estudo piloto pesquisou as redações dos alunos de 6º ano/B30 das duas escolas (todos sobre a mesma proposta), servindo como uma espécie de contraponto em relação aos textos do 8º ano, totalizando 30 textos por escola.

Os procedimentos e resultados desses estudos, bem como as variações ocorridas de um estudo para o outro, conforme fomos aumentando e diversificando o *corpus* de estudo, são apresentados nas próximas seções.

### 5.3.1 Estudo Piloto I

O primeiro estudo piloto foi realizado com apenas oito textos (quatro de cada escola), que representavam 20% do total de textos da única proposta com a mesma temática nas duas escolas: “A vida como filme”. Esses quatro textos foram escolhidos de maneira que, mesmo heterogêneos, fossem equivalentes entre as duas escolas, tanto em número de linhas, quanto nas avaliações realizadas pela professora e notas/conceitos atribuídos aos textos. A amostra procurou regularidade nesses sentidos e foi constituída dessa forma:

Quadro 13 – Textos do primeiro estudo piloto

Escola Municipal	Escola Estadual
TURMAS C21 e C22	TURMA 81
4 textos	4 textos

Fonte: Elaboração própria.

Em relação aos critérios, os oito textos desse primeiro estudo foram constituídos de:

Tabela 2 – Redação por escola (primeiro estudo piloto) – 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: Elaboração própria.

Vale ressaltar que as avaliações são diferentes nos dois sistemas de ensino público. Enquanto as escolas municipais utilizam conceitos, que variam de A (atingiu os objetivos) até NA (não atingiu os objetivos), as escolas estaduais dispõem de avaliação numérica. Portanto, a mesma atividade de produção textual realizada nas duas escolas teve valor máximo de A na escola municipal e, na escola estadual, o valor máximo foi de 3,0 pontos (que, juntamente com outras atividades avaliativas, somadas, formaram a nota final do aluno na disciplina de Língua Portuguesa daquele trimestre). Sendo assim, analisamos, nesse primeiro estudo, pouco mais de 100 linhas de texto escrito de cada escola, todos com avaliações semelhantes, independente do sistema avaliativo da rede/escola.

Como contraste aos textos dos alunos, utilizamos, nesse primeiro estudo, o texto de apoio à produção da redação, de Scliar (2003) (ANEXO III) e um texto do livro didático de cada escola (ANEXO IV), lido e interpretado pelos alunos antes da atividade de produção. Dessa forma, a composição geral desse primeiro estudo piloto foi composta conforme o Quadro 14 abaixo:

Quadro 14 – Composição geral do *corpus* do estudo piloto I

	<b>Número de redações</b>	<b>Texto de apoio</b>	<b>Textos – livros didáticos</b>
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: Elaboração própria.

Os textos de contraste foram também processados no *software AntConc*, e deles foram considerados os mesmos elementos observados nas redações. Assim, conseguimos comparar esses materiais. Conforme Silva (2020, p. 38):

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.

O primeiro resultado obtido foi o número de *tokens* e, depois, o número de *types*, conforme a Tabela 3 a seguir:

Tabela 3 – Quantidade *tokens/types* – estudo piloto I

	<b>Redações do Grupo A (Escola Municipal)</b>	<b>Redações do Grupo B (Escola Estadual)</b>	<b>Texto de Apoio (Proposta do tema)</b>	<b>Texto do livro A</b>	<b>Texto do livro B</b>
<i>Tokens</i>	836	729	409	547	699
<i>Types</i>	350	306	214	334	346

Fonte: Elaboração própria.

Como é possível visualizar na Tabela 3, o número total de palavras escritas nos textos dos alunos foi de 836 na escola municipal, frente a 729 na escola estadual, uma diferença de 107 palavras a mais no total escrito pelos educandos da primeira escola.

A média de palavras por texto foi de 209 na escola municipal e de 182 na escola estadual. Essa média nos dá o número aproximado de palavras que cada um dos alunos escreveu nos seus textos. Nesse sentido, é possível fazer comparações com os materiais de apoio, visto que cada um deles representa também um único texto. Percebemos, então, que essa média em muito difere dos resultados do texto de apoio e dos textos dos livros didáticos, que apresentaram 409, 547 e 699 palavras por texto, respectivamente, ou seja, o dobro ou até mais palavras que nos textos dos estudantes.

Em relação ao número de *types*, os textos da escola municipal apresentaram 350 palavras diferentes não repetidas, enquanto os da escola estadual apresentaram 306. A diferença, nesse caso, foi menor de uma escola para outra: apenas 44 *types* a mais na escola do estado, considerando os quatro textos de cada grupo. Também não houve grande variação de *types* entre os textos dos livros didáticos, com 344 e 346 *types*, respectivamente.

Em relação aos graus de frequência e à proporção de emprego de palavras lexicais e gramaticais, os textos analisados seguiram o padrão da Língua Portuguesa, de apresentar palavras gramaticais em maior quantidade e com mais frequência, como é possível observar na Tabela 3. A Tabela 4 apresenta o recorte



das vinte primeiras palavras mais frequentes no conjunto dos quatro textos dos alunos de cada escola.

Tabela 4 – 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: Elaboração própria.

Além das palavras gramaticais, as que se repetiram nos dois grupos, considerando a base semântica das palavras, foram as grifadas em azul na Tabela 4: filme, vida, pessoas e o verbo ser. Tais palavras podem ser também chamadas de palavras-tópico ou tema da proposta de produção. Em função disso, também apareceram como mais frequentes no texto de apoio. Entretanto, não aparecem nos textos dos livros didáticos, uma vez que tais textos não se referem ao mesmo assunto. Isso é o que mostra a Tabela 5:

Tabela 5 – 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. Roxo – palavras lexicais somente de um grupo.

Fonte: Elaboração própria.

Na Tabela 5, das primeiras vinte palavras mais frequentes dos textos de contraste, observamos que as gramaticais seguem como as mais frequentes e que, entre as lexicais, somente filme e vida se repetem no texto de apoio e nos textos dos alunos. O verbo ser também aparece nos textos dos alunos conjugado no tempo passado como seria, enquanto aparece nos textos dos livros didáticos conjugado no tempo presente como é. Esse verbo também é um dos mais frequentes na língua como um todo.

Sendo assim, podemos dizer que, muito provavelmente, os alunos somente ativaram suas memórias curtas aos escreverem seus textos, pois aproveitaram o vocabulário do texto de apoio. Por outro lado, aparentemente, não ativaram suas memórias longas, visto que o vocabulário dos textos didáticos não foi verificado nas redações. Somente as palavras gramaticais são comuns entre os textos dos livros didáticos e os textos dos alunos. Isso confirma um padrão de normalidade gramatical do uso da língua escrita e reafirma o achado de Biderman (1998a), que acredita em um possível núcleo de vocabulário, que seria usado em qualquer tipo de texto, em nossa língua.

A riqueza lexical, medida que relaciona o número de *tokens* e de *types* de um texto, também foi analisada, conforme Tabela 6, abaixo:

Tabela 6 – 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: Elaboração própria.

Essa proporção, que mostra o quanto um texto é mais ou menos variado em relação ao vocabulário empregado, não variou de uma escola para outra, ficando em torno de 42% em ambas. Essa medida se aproxima à riqueza de um texto jornalístico, que busca seguir o preceito da boa redação jornalística de não repetir palavras e tem riqueza em torno de 40% (BISOGNIN, 2009).

Em relação aos textos de contraste, houve variação e aumento dessa medida no texto de apoio e no texto do livro didático da escola municipal. Todavia, a riqueza lexical dos textos dos alunos se aproxima da riqueza do texto do livro didático da escola estadual, conforme observamos na Tabela 6, em torno de 41%.

Por fim, desses mesmos textos obtivemos os resultados dos itens *hapax legomena*, de frequência única, pois, segundo Silva (2020, p. 43):

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 sentido, as palavras presentes nesse grupo de frequência única espelham a particularidade do *corpus* analisado. Assim, para analisar o *hapax* dos *corpora*, fizemos o recorte das palavras que ocorrem apenas uma vez no *corpus* e iniciam pela vogal A e pela consoante P.

A escolha pelos itens iniciados por essas letras se justifica com base em estudos de Lexicografia, que exploram a cobertura de repertório vocabular em dicionários por meio de amostragens. Tais letras tendem a reproduzir um espectro variado de classes de palavras – artigos, substantivos, adjetivos e verbos. Essa

técnica de amostragem é encontrada em trabalhos de Damim (2005) e em materiais do Programa Nacional do Livro e do Material Didático (PNLD) sobre avaliação de Dicionários (BRASIL, 2012). Dessa maneira, para o nosso objetivo, foi importante selecionar letras que abrangessem a maior quantidade e diversidade de itens em relação ao repertório vocabular observado.

Na Tabela 7 visualizamos o recorte da vogal A, referente aos textos dos dois grupos de alunos, e contrastamos com os mesmos dados referentes ao texto de apoio e aos textos didáticos:

Tabela 7 – Amostra de *hapax legomena* dos *corpora* analisados – vogal A

	GRUPO A (Escola Municipal)	GRUPO B (Escola Estadual)	Texto de Apoio (A vida como filme)	Texto didático prefeitura	Texto didático estado
1.	acabar	<b>aconteceria</b>	aborrecidos	abdicar	acabado
2.	<b>acontece</b>	<b>acontecido</b>	acesso	absolutamente	acertos
3.	<b>aconteceria</b>	admiro	agradecendo	acessam	<b>acontecendo</b>
4.	<b>aconteceu</b>	<b>adolescente</b>	ah	acontecem	<b>alguma</b>
5.	<b>adolescencia</b>	adulta	ainda	<b>adolescente</b>	<b>algumas</b>
6.	<b>agora</b>	agente	ali	ainda	<b>alguns</b>
7.	ajeita	<b>agora</b>	alternativo	<b>algum</b>	aliás
8.	<b>algo</b>	agradecer	amigos	além	alta
9.	<b>algum</b>	agradeço	amplos	amiga	altura
10.	<b>algumas</b>	ajudou	<b>amáveis</b>	antes	antes
11.	<b>alguns</b>	alexandre	aparecia	antigamente	aparte
12.	<b>almoço</b>	<b>algumas</b>	apenas	antigo	apertando
13.	<b>amadas</b>	<b>almocei</b>	aplausos	<b>anuidade</b>	aproveitou
14.	<b>amo</b>	alta	apresentador	aparelho	assim
15.	<b>ano</b>	amiga	aquilo	aparelhos	assunto
16.	<b>ao</b>	amigas	ar	aposentadoria	atenção
17.	apesar	<b>amor</b>	as	assim	atenções
18.	aquelas	<b>anos</b>		assinatura	atingir
19.	assim	<b>ao</b>		assustou	aula
20.	ate	apaixonada		<b>ator</b>	aí
21.	<b>ator</b>	apaixonaria		através	
22.	atuar	araujo		aumentaram	
23.	aturar	assistir		azul	
24.	aula	<b>atris</b>			
25.	<b>avó</b>	<b>avós</b>			
26.	azar				
27.	ações				
28.					
29.					
30.					
31.					

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

Conforme a Tabela 7, considerando a base semântica das palavras, há 52% de *hapax* em comum entre os textos dos alunos das duas escolas, o que demonstra aproximação do vocabulário específico desses dois grupos de estudantes. Em

relação aos textos de contraste, percebemos que há pouca semelhança. O texto de apoio apresenta apenas uma palavra semanticamente semelhante: amáveis. Já os textos didáticos apresentam três e quatro coincidências semânticas, respectivamente, o que demonstra um vocabulário diferenciado em relação àquele utilizado pelos alunos.

A Tabela 8 apresenta o mesmo recorte anterior, mas com as palavras do *hapax* iniciadas pela consoante P:

Tabela 8 – Amostra de hapax legomena dos corpora analisados – consoante P

	GRUPO A (Escola Municipal)	GRUPO B (Escola Estadual)	Texto de Apoio (A vida como filme)	Texto didático Prefeitura	Texto didático Estado
1.	palhaçada	paeva	país	pacotes	papel
2.	parecidos	pagaram	palco	parava	parabéns
3.	participar	parte	papel	parecia	parece
4.	paz	partes	pessoas	parte	parede
5.	pela	partiria	planejamento	particular	parou
6.	pelos	passa	plateia	passada	parte
7.	perfeita	pegaria	<b>pode</b>	passado	participação
8.	pessoa	pensei	<b>poderíamos</b>	país	passados
9.	piscina	perdido	preferimos	pela	passos
10.	possíveis	perigosas	prima	pelo	pedira
11.	praça	preferido	princesa	perder	peito
12.	peça	<b>primeira</b>	proporcionar	perderam	pensou
13.	prima	<b>primeiro</b>	prêmio	permanecer	pequena
14.	<b>primeiramente</b>	professoras	pudéssemos	pessoas	perguntando
15.	<b>primeiro</b>	proposta		pet	pescou
16.	pro	protornizado		pior	<b>podia</b>
17.	<b>pudecem</b>	<b>pudesse</b>		plano	pouco
18.	<b>pudese</b>			<b>pode</b>	pra
19.	público			<b>podem</b>	precisava
20.				<b>poder</b>	prestar
21.				preciso	<b>primeira</b>
22.				preço	principais
23.				<b>primeira</b>	principal
24.				produto	pé
25.				produtos	
26.				profissionais	
27.				proporção	
28.				puro	
29.				pão	
30.				pé	
31.				públicos	

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

No universo vocabular da letra P, como podemos observar por meio da Tabela 8, apenas 19,5% das palavras coincidem semanticamente entre os grupos das duas escolas. Somente as palavras: “primeiramente”, “primeiro”, “pudecem”, “pudese”, “primeira”, “primeiro” e “pudesse” foram consideradas semanticamente

semelhantes. Em relação ao texto de apoio, somente a base semântica do verbo “poder” foi coincidente e, nos textos didáticos, o mesmo verbo “poder” e o numeral “primeira”.

A partir do recorte dessas duas letras, percebemos que há maior coincidência no *hapax* da letra A, que também apresenta maior quantidade de palavras. Além disso, foi possível observar que não há quantidade de palavras significativas que coincidam entre os textos dos alunos e os materiais de contraste retirados dos livros didáticos, o que já havia sido salientado na análise das palavras mais frequentes.

Em suma, a partir do primeiro estudo piloto realizado, concluímos, principalmente, que os grupos não apresentam muita diferença em termos de vocabulário escrito, palavras mais usadas e riqueza lexical. Além disso, conforme Tcacenco, Silva e Finatto (2020, p. 216-217):

A partir da análise da amostra, é possível verificar, preliminarmente, que:

- a) os textos escolares, assim como os demais textos sob exame, seguem um padrão de “normalidade” quanto à distribuição e frequência de palavras gramaticais;
- b) a primeira palavra lexical mais frequente nas redações e no texto da proposta também está no padrão de “normalidade”, dado que “filme” corresponde ao tópico dos textos e ao da proposta de produção;
- [...]
- d) ao escreverem seus textos, aparentemente, os estudantes ativaram, sobretudo, suas memórias curtas, já que o tema principal do texto de apoio e da proposta de produção textual (“filme”) e as palavras que o rodeiam (“vida”, “seria”, “não”) coincidem;
- e) não há coincidência entre o repertório de palavras da categoria *hapax legomena* entre as redações e os textos dos livros didáticos a que os estudantes foram expostos antes de realizarem as suas produções.

Então, esse primeiro estudo piloto serviu para reavaliarmos algumas questões, a fim de projetar a pesquisa em grande escala, como:

- a) ponderar o melhor percentual de amostra para estudo em grande escala, pois aumentando o número de redações os resultados podem vir a ser diferentes;

- b) administrar a heterogeneidade de dados que compõem o *corpus*, já que os temas dos textos são variados, com propostas diferentes de uma escola para outra, e isso pode interferir nos resultados;

- c) mensurar o papel do texto do livro didático em relação ao vocabulário dos alunos, pois, pelo que percebemos inicialmente, o texto do livro didático não influencia no vocabulário escrito das redações.

Acreditamos, também, que o primeiro estudo levou à reflexão sobre aquela percepção inicial, percebida também por colegas de outras áreas, de que os alunos da rede municipal teriam mais dificuldades de aprendizagem, e que as produções escritas, por meio do vocabulário, espelhariam essas questões. A partir dos resultados deste estudo inicial, talvez, na prática, as questões não sejam exatamente como pensamos, pelo menos não do ponto de vista léxico-estatístico. Isso instigou-nos a realizar o segundo estudo, com todo o conjunto de dezoito textos (de cada escola) sobre a mesma proposta de produção: “A vida como um filme”. O segundo estudo é descrito na próxima seção.

### **5.3.2 Estudo Piloto II**

O estudo piloto II reuniu todos os textos referentes à proposta: “A vida como filme”. Ele serviu como uma espécie de continuidade do primeiro estudo, já que as redações nele estudadas foram sobre a mesma temática. A ideia principal foi observar se os resultados persistem à medida que aumentamos o número de produções analisadas ou não. Dessa forma, utilizamos o total de textos disponíveis dessa proposta: dezoito textos de cada escola. Esse foi o número de redações que atenderam à proposta, tanto em relação ao número de linhas, como em relação ao tema proposto.

O texto de apoio, “A vida como filme”, de Moacyr Scliar (2003), continuou como fonte de contraste, visto que, conforme apontado pelo primeiro estudo, os alunos utilizaram com frequência as palavras-tópico da proposta, presentes também em grande número no texto de apoio.

Como, por meio do primeiro estudo piloto, percebemos que os textos dos livros didáticos não se mostraram satisfatórios como contraste ao vocabulário escrito dos alunos, uma vez que eles não utilizaram o mesmo vocabulário escrito desses textos nas suas redações, optamos por não os utilizar nesse segundo estudo. Então, nosso *corpus* do segundo estudo piloto ficou organizado conforme o Quadro 15 a seguir:

Quadro 15 – Composição geral do *corpus* do estudo piloto II

	<b>Número de redações</b>	<b>Texto de Apoio</b>
Grupo A (Escola Municipal)	18 textos	1: A vida como filme, de Moacyr Scliar
Grupo B (Escola Estadual)	18 textos	1: A vida como filme, de Moacyr Scliar

Fonte: Elaboração própria.

Desse material, os primeiros resultados obtidos foram o número total de palavras e o número de palavras diferentes repetidas, conforme mostra a Tabela 9:

Tabela 9 – Quantidade *tokens/types* – estudo piloto II

	<b>Texto de apoio: A vida como filme</b>	<b>18 textos – Escola Municipal</b>	<b>18 textos – Escola Estadual</b>
<i>Tokens</i>	409	3867 (média de 215/texto)	3902 (média de 217/texto)
<i>Types</i>	214	1076 (média de 120/texto)	1167 (média de 121/texto)

Fonte: Elaboração própria.

O número total de palavras não variou muito de uma escola para a outra, 3.867 palavras nos textos da escola municipal e 3.902 palavras nos textos da escola estadual. Quando analisamos a média de palavras por texto, percebemos que, de fato, essa diferença realmente não é considerável, pois varia de 215 para 217 palavras por texto, o que aproxima os grupos nesse quesito.

Entretanto, ao analisarmos o número total de palavras em relação ao texto de apoio, que é um só e tem 409 palavras, frente a um universo de cerca de 216 palavras por texto nos textos dos alunos, aí encontramos uma variação expressiva. Os alunos escrevem um pouco mais do que a metade do texto em que se basearam, no quesito quantidade de palavras. Acreditamos que isso pode estar atrelado às propostas de produção (ANEXO II), com número mínimo de linhas, bem como à efemeridade e ao imediatismo da vida adolescente nos dias de hoje.

Já em relação ao número de palavras diferentes que não se repetem, a diferença de uma escola para a outra é em torno de 90 palavras diferentes a mais nas redações da escola do estado, que também apresenta maior quantidade no total de palavras. Ao compararmos a média de *types* entre os dezoito textos dos alunos (que também foi muito próxima, cerca de 120 *types* por texto) e o texto de apoio, temos também diferença expressiva, pois as produções dos alunos, em média, exibem 120 *types* frente às 214 do texto de apoio.



Em relação à frequência e ao emprego de palavras gramaticais e lexicais, podemos ver, na Tabela 10, que não houve grande variação de uma escola para a outra.

Tabela 10 – Relação de palavras lexicais e gramaticais – estudo piloto II

	Escola Municipal	Escola Estadual
Palavras Gramaticais	<b>e, que, eu, a, de, minha, o, para, um, não, no, com, as, meu, mais, uma, se, na, me, os, do, muito, depois, quando, em, da, meus, ela, pra, até, ele, tudo, hoje, mas, por, ou, porque</b> (total: 37 / 74%)	<b>e, que, eu, a, o, de, um, minha, para, uma, meu, não, no, com, muito, por, mas, em, se, do, da, as, mais, me, como, na, meus, eles, os, sua, quando, até, ele, depois, seu</b> (total: 35 / 70%)
Palavras Lexicais	<b>filme, dia, vida, casa, seria, mãe, pessoas, amigos, coisas, iria, era, escola, ia</b> (total: 13 / 26%)	<b>filme, seria, vida, é, foi, final, era, pessoas, dia, mãe, mauro, amigos, fazer, ser, tem</b> (total: 15 / 30%)

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

Na Tabela 10 estão as 50 palavras mais frequentes, considerando todos os dezoito textos sobre a mesma proposta das duas escolas, separadas entre palavras gramaticais e palavras lexicais. Como é possível visualizar, cerca de 70%, nos dois grupos, são palavras gramaticais, que são as de maior frequência na língua como um todo.

Entre as palavras lexicais, temos coincidência semântica de apenas oito palavras, grifadas na Tabela 10 em negrito: o verbo mais usado no português como um todo, “ser”, e as palavras que são os temas da proposta: “filme” e “vida”. As outras palavras lexicais em comum nos dois grupos foram: “pessoas”, “dia”, “mãe” e “amigos”. Essas palavras, apesar de não estarem na proposta, se repetiram em ambos os grupos, demonstrando aproximação de assuntos de preferência desses jovens ao escrevem suas produções.

“Filme”, “vida” e o verbo “ser”, por serem palavras-tópico, aparecem também em grande quantidade no texto de apoio, como mostra o Quadro 16:

Quadro 16 – Palavras lexicais e gramaticais em relação ao texto de apoio – estudo piloto II

	Texto de Apoio	Escola Municipal	Escola Estadual
Palavras Gramaticais	<b>que, um, de, o, e, uma, nosso, não, a, ao, do, para, em, os, se, com, na, nossa, ou, aos, ela</b> , isso, já, no, nos, nossos, aquelas, <b>até</b> , da, dessas, grande, <b>mais, mas</b> , menos, <b>muito</b> , nem	<b>e, que, eu, a, de, minha, o, para, um, não, no, com, as, meu, mais, uma, se, na, me, os, do, muito, depois, quando, em, da, meus</b> , ela, pra, <b>até, ele</b> , tudo, hoje, <b>mas</b> , por, ou, porque	<b>e, que, eu, a, o, de, um, minha, para, uma, meu, não, no, com, muito</b> , por, <b>mas, em, se, do, da, as, mais, me</b> , como, <b>na, meus</b> , eles, <b>os</b> , sua, <b>quando, até, ele, depois</b> , seu
Palavras Lexicais	<b>filme, vida</b> , história, tem , estaríamos, oscar, <b>é</b> , atores, atrizes, coisa, fazer, final, momento, poderes	<b>filme</b> , dia, <b>vida</b> , casa, <b>seria</b> , mãe, pessoas, amigos, coisas, iria, <b>era</b> , escola, ia	<b>filme, seria, vida, é</b> , foi, final, <b>era</b> , pessoas, dia, mãe, mauro, amigos, fazer, ser, tem

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

Em relação aos textos de apoio, o padrão segue o mesmo, pois vemos as palavras gramaticais como as mais frequentes e em maior quantidade também. E, entre as lexicais, as palavras-tópico da proposta, que se repetem nas duas escolas, também se repetem no texto de apoio: “filme”, “vida” e o verbo “ser”, nas suas diferentes conjugações.

É importante salientar que, a partir do Quadro 16, percebemos que algumas das palavras lexicais coincidentes nos dois grupos de alunos não apareceram entre as mais frequentes do texto de apoio: “dia”, “mãe”, “amigos” e “pessoas”. Tais palavras fizeram parte do vocabulário escrito mais frequente nos textos dos alunos, mesmo sem ter como fonte o texto de apoio. Além disso, a coincidência desse repertório usado com preferência aproxima os perfis vocabulares desses dois grupos.

Outra medida analisada foi a riqueza lexical, conforme a Tabela 11:

Tabela 11 – Riqueza lexical – estudo piloto II

	Texto de Apoio	Escola Municipal	Escola Estadual
Riqueza Lexical	52,3%	27,8%	29,9%

Fonte: Elaboração própria.

A riqueza lexical praticamente não variou entre os grupos de textos das duas escolas, pois ficou em torno de 28% e 30%, que é uma riqueza semelhante a de um artigo científico da área de biologia, em que geralmente se repetem os termos da área como forma de reiterar o assunto, ficando em torno de 29% (FINATTO; CREMONESE; AZEREDO, 2008). Essa medida variou em relação ao primeiro

estudo, com apenas quatro textos de cada grupo, pois a riqueza era de aproximadamente 42%. Quando se analisa um número menor de textos, a tendência é que essa riqueza de fato seja maior, uma vez que o número total de palavras é menor, então é natural que haja maior variação nesse vocabulário escrito. Todavia, à medida que se aumenta o número de textos e, conseqüentemente, o número total de palavras, essa riqueza tende a diminuir, uma vez que o espectro do vocabulário fica mais limitado e, por consequência, repetitivo.

A riqueza lexical do texto de apoio segue em 52% e, nesse sentido, difere mais dos textos dos alunos do que no primeiro estudo, com média de 29%.

Em relação ao universo de palavras de frequência 1, o *hapax*, o nosso recorte foi novamente das palavras que iniciam com a vogal A e com a consoante P. Na Tabela 12, abaixo, temos como exemplo as primeiras quinze palavras do *hapax* de cada grupo iniciando pelas letras selecionadas:

Tabela 12 – *Hapax legomena* – estudo piloto II

	Escola Municipal	Escola Estadual
Letra A – quantidade	67	76
Letra A – exemplos	abc, aborta, <b>abri, acaba, acabar, acabou, acada, achei, aconteceria, aconteceriam, acontecido</b> , acordava, acre, adaptando, adoeceu...	<b>abriu, acabar, achei, acontecem, acontecer, aconteceram, acontecido, acontecimentos, aconteça</b> , acostumar, acreditava, admiro, <b>adolescência, adolescente, adolescência...</b>
Letra P – quantidade	77	82
Letra P – exemplos	padaria, <b>pagar</b> , paisagens, palhaçada, palmares, papel, paraná, <b>parar, parecia, parecidos</b> , parede, participar, participaram, <b>partiu</b> , parto...	padrão, paeva, <b>pagaram, parados, parece, parecem</b> , partes, <b>partir, partiria, passa, passam, passar, passaria</b> , patos, patoscidym...

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

A partir do recorte do *hapax* dessas duas letras, observamos que, iniciando pela vogal A, a quantidade é de 67 palavras num grupo e de 76 no outro. Percebemos, pelos exemplos grifados em negrito, que, considerando a base semântica das palavras dos exemplos da Tabela 12, cerca de 67% delas se repetem.

Já com a consoante P temos maior número total de palavras repetidas uma única vez nos dois grupos, 77 e 82, respectivamente. Como é possível visualizar na

Tabela 12, por meio das palavras em negrito, 50% das palavras do *hapax* com inicial P, dentro dos nossos exemplos, se repetem nos dois grupos.

A partir desse segundo estudo piloto, também constatamos alguns aspectos que precisavam ser reavaliados e repensados para a pesquisa em maior escala, com todo nosso *corpus* de estudo. O primeiro deles segue sendo a heterogeneidade de dados que compõem esse *corpus*. Como as temáticas das produções escritas dos alunos são variadas, é preciso mensurar o quanto isso pode interferir nos resultados ao serem misturadas todas as temáticas disponíveis e ao analisar todo o *corpus* de estudo do 8º ano/C20 (que é onde estão esses temas diversos). Isso é realizado no estudo piloto III. Esse parece ser o maior desafio da pesquisa como um todo.

Outro aspecto necessário de ser reavaliado é o do tratamento lexical estatístico, pois precisa ser melhorado. Percebemos, conforme os dados desse segundo estudo piloto, muitas variedades ortográficas de um mesmo item (*escrevem* com s, ss, z), então, é preciso repensar se tais palavras podem ser consideradas dois itens diferentes, já que os escritores as utilizaram com o mesmo sentido. O mesmo acontece com os verbos e suas conjugações possíveis, assim como com substantivos e adjetivos e suas variantes de gênero e número. Muitos deles aparecem como *hapax* por não se encontrarem lematizados, o que, além de não deixar essa camada tão específica, também aumenta a quantidade dessas palavras de frequência única. Nesse sentido, acreditamos que, lematizando as contagens, o nosso *hapax* ficaria mais reduzido e espelharia melhor as especificidades de cada grupo. Assim, seria possível de fato contrastar se as palavras específicas de cada grupo coincidem ou não e em que medida.

O terceiro estudo piloto, que é apresentado a seguir, utilizou como *corpus* principal todo o conjunto de 94 textos dos alunos do 8º ano/C20 das duas escolas.

### 5.3.3 Estudo Piloto III

O terceiro estudo piloto utilizou como fonte principal o total de textos dos alunos do 8º ano/C20 disponíveis no acervo pessoal da professora: 94 redações, sendo 47 produções de cada escola. Esse total de textos contém cinco temáticas diferentes, duas relacionadas à interpretação de filmes (“Um sonho possível” e “Os

Goonies”) (ANEXO V), duas com textos de apoio (“Voltar de férias é...” e “Futuro”) (ANEXO III), e aquela única proposta que se repete nas duas escolas, que também tem texto de apoio e que foi analisada nos dois primeiros estudos piloto (“A vida como filme”) (ANEXO III). Os textos do estudo piloto III estão organizados conforme o Quadro 17:

Quadro 17 – *Corpus* de estudo – estudo piloto III

<b>Escola Municipal</b>	Proposta	Quantidade	<b>Escola Estadual</b>	Proposta	Quantidade
	1: A vida como filme	18 textos		1: A vida como filme	18 textos
	2: Um sonho possível	15 textos		4: Os Goonies	14 textos
	3: Voltar de férias é...	14 textos		5: O Brasil que eu quero	15 textos
Geral/total	3 propostas (1,2,3)	47 textos	Geral/total	3 propostas (1, 4, 5)	47 textos

Fonte: Elaboração própria.

Além do *corpus* de estudo, com todos os textos dos alunos de 8º ano, foram usados como contraste, neste terceiro estudo, os textos de apoio às propostas de produção, quando existentes. Das cinco propostas, três contaram com texto de apoio, de maneira que os textos usados como contraste foram os seguintes, expostos no Quadro 18:

Quadro 18 – Textos de contraste – estudo piloto III

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
Proposta/Texto de apoio	1: A vida como filme – Moacyr Scliar (2003)	1: A vida como filme – Moacyr Scliar (2003)
	3: Voltar de férias é... – Nathália Prestes Bosak (2012)	5: Futuro – Nilson Souza (2018)

Fonte: Elaboração própria.

Neste terceiro estudo piloto, como as temáticas eram variadas e cada texto de apoio se relacionava com uma quantidade específica de redações dos alunos, os dados obtidos foram apresentados como um todo, mas os contrastes foram realizados de maneira pontual. Ou seja, para cada item analisado, foram expostos

os dados do conjunto total de textos, separados somente por escola. Entretanto, quando realizado o contraste com os textos de apoio, esses dados foram separados por proposta de produção, para que fosse possível mensurar em que medida os alunos, aos escreverem suas redações, se aproximaram ou se distanciaram dos textos de apoio.

O primeiro resultado obtido foi o número total de palavras, como mostra a Tabela 13:

Tabela 13 – *Tokens/types* – estudo piloto III

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
<i>Tokens</i>	8382 (média de 178,3/texto)	9925 (média de 211,1/texto)
<i>Types</i>	1731	2196

Fonte: Elaboração própria.

O número total de palavras foi de 8.382 na escola do município, frente a 9.925 na escola do estado, cerca de 1.500 palavras a mais na escola estadual. A média de palavras por texto foi de 178 na primeira escola e de 211 na segunda, ou seja, os alunos da escola estadual escreveram aproximadamente 23 palavras a mais que os educandos da escola municipal em cada texto. Então, nesse conjunto maior, percebemos que a quantidade de palavras (total e média por texto) já se diferenciou mais entre os grupos do que nos estudos piloto anteriores.

Já em relação ao número de *types*, a diferença é menor de uma escola para a outra, em torno de 460 palavras. A média de palavras diferentes na escola municipal foi de cerca de 103 *types* por texto, enquanto a média da escola estadual foi de 118 *types* por texto. Nesse sentido, constatamos que também houve maior diferença em relação à quantidade de palavras diferentes quando analisamos um conjunto maior de redações.

Em relação aos textos de apoio, o primeiro, “A vida como filme”, relativo à mesma proposta nas duas escolas, tem praticamente o dobro de palavras que os textos dos alunos. Como isso já foi observado no estudo piloto anterior, optamos por não repetir as observações aqui.

Outro texto de apoio contrastado foi “Voltar de férias é...” cujos dados estão na Tabela 14:

Tabela 14 – *Tokens/types* – proposta 03

	<b>Texto de Apoio – Voltar de férias é...</b>	<b>Textos da proposta 03 – Escola Municipal</b>
<i>Tokens</i>	357	2101 (média de 150/texto)
<i>Types</i>	187	572 (93,2/texto)

Fonte: Elaboração própria.

Nessa proposta, escrita somente pelos alunos da escola municipal, a quantidade de palavras variou ainda mais, visto que a média dos textos dos alunos foi de 150 palavras por texto, frente às 357 palavras escritas no texto de apoio. Nesse sentido, o texto de apoio apresentou mais que o dobro das palavras escritas pelos alunos nas suas redações. Em relação ao número de *types*, a média dos textos dos educandos foi de 93 *types* por texto, frente às 187 palavras do texto de apoio. Isso mostrou que o vocabulário que os estudantes exibiram nas suas produções tem praticamente a metade das palavras diferentes que o texto de apoio no qual se basearam para a sua escrita.

O último texto de apoio contrastado foi “Futuro”, trabalhado somente com os alunos da escola estadual. Os dados desse contraste são apresentados na Tabela 15:

Tabela 15 – *Tokens/types* – proposta 05

	<b>Texto de Apoio – Futuro</b>	<b>Textos da proposta 05 – Escola Estadual</b>
<i>Tokens</i>	431	2230 (média de 149/texto)
<i>Types</i>	256	712 (média de 91/texto)

Fonte: Elaboração própria.

Nesse contraste, a diferença em quantidade de palavras foi ainda mais relevante, como é possível visualizar na Tabela 15. Foram 431 palavras no texto de apoio, frente à média de 149 palavras nos textos dos alunos. Isso representa quase duas vezes mais palavras no texto de apoio. Os mesmos números quantitativamente diferentes apareceram em relação às *types*. Enquanto a média nas produções dos alunos é de 91 palavras diferentes por redação, o texto de apoio apresentou 256 *types*. Ou seja, havia quase duas vezes mais *types* no texto de apoio.

Quanto aos textos de apoio, então, que existiam em 3 propostas, percebemos que esses sempre apresentam mais quantidade de palavras que a média dos textos dos alunos. Isso também acontece quanto ao número de palavras diferentes, que sempre foi expressivamente maior nos textos de apoio.

Sabemos que não podemos comparar, no mesmo patamar, o texto de um escritor e o texto de um aluno. Porém, o segundo texto de apoio, “Voltar de férias é...”, na época em que foi publicado no jornal, foi escrito também por uma estudante. Dessa forma, foi importante o contraste com esse texto especificamente, pois seus redatores de alguma forma se aproximam do contexto da autora. Ainda assim, percebemos que os resultados de *tokens* e de *types* seguem distantes, mesmo nessa comparação.

Quanto aos graus de frequência e relação de palavras lexicais e gramaticais, percebemos que com maior quantidade de textos o padrão da língua como um todo permanece: palavras gramaticais como as mais frequentes e palavras lexicais típico da proposta também aparecem com alta frequência. Isso é o que foi observado na Tabela 16:

Tabela 16 – 50 primeiras palavras mais frequentes – estudo piloto III

	Escola Municipal	Escola Estadual
Palavras Gramaticais	<b>e, que, de, a, o, não, eu, para, um, as, com, os, se, minha, no, mais, na, uma, ele, do, por, ou, em, da, quando, meu, muito, depois, me, pra, tudo, você, todos, mas, até</b> (total: 35 / 70%)	<b>e, que, o, de, a, um, eu, os, para, não, eles, do, uma, com, mais, no, em, por, se, mas, as, da, minha, muito, na, ele, meu, depois, quando, sua, como, então, seu, todos, me, sem, até</b> (total: 37 / 74%)
Palavras Lexicais	<b>é, filme, vida, dia, pessoas, escola, ter, ser, voltar, amigos, bom, casa, tem, coisas, ano</b> (total: 15 / 30%)	<b>filme, é, brasil, casa, pessoas, tem, seria, tesouro, foi, amigos, era, quero, vida</b> (total: 13 / 26%)

Fonte: Elaboração própria.

Conforme a Tabela 16, podemos perceber que não houve variação de uma escola para a outra entre as 50 palavras mais frequentes, considerando os 47 textos de cada escola, com diferentes propostas, misturados entre si. Cerca de 70%, dentro desse recorte de exemplo, são palavras gramaticais e, entre as lexicais, há coincidência de dez palavras. Entre elas estão os verbos que geralmente são mais usados no português: “ser” e “ter”, e as palavras que são os temas da única proposta desse conjunto de textos que coincide nas duas escolas: “filme” e “vida”. Algumas outras palavras também coincidem nos dois grupos: “casa”, “amigos”, “pessoas”.

O contraste entre os textos dos alunos e o texto de apoio “A vida como filme”, já foi apresentação nas seções anteriores, portanto, optamos por não o repetir aqui.



Na Figura 5, é possível visualizar os contrastes entre os outros textos de apoio e os textos dos alunos referentes àquela proposta específica:

Figura 5 – Palavras frequentes em relação aos textos de apoio

	Texto de Apoio – Voltar de férias é...	Textos da proposta 03 – Escola Municipal
Palavras Gramaticais	<b>de, que, a, o, e, com,</b> da, <b>para,</b> em, <b>eu, as,</b> do, <b>não, os,</b> depois, <b>na,</b> primeiro, <b>se,</b> sem	<b>e, que, de, a, os, para,</b> <b>as, o, não, eu,</b> mais, <b>se, na, com,</b> no, um
Palavras Lexicais	<b>é, aula, dia,</b> retomar, ir, férias, ir	<b>é, voltar, aulas,</b> ano, escola, amigos, <b>aula,</b> professores, <b>dia</b>

	Texto de Apoio – O Brasil que eu quero	Textos da proposta 05 – Escola Estadual
Palavras Gramaticais	<b>de, que, o, e, a, para,</b> <b>não,</b> do, <b>em,</b> na, <b>um,</b> como, <b>com,</b> da, mas, meu, nem, nossa, todos	<b>que, e, o, para, não,</b> eu,as, <b>um,</b> mais, de, a, os, <b>com,</b> sem, se, isso, <b>em,</b> por, muito
Palavras Lexicais	<b>é, escola, futuro,</b> <b>brasil,</b> era, horizontal	<b>brasil,</b> quero, pessoas, <b>é, tem, futuro, escolas</b>

Fonte: Elaboração própria.

Em relação ao segundo texto de apoio, as palavras lexicais que coincidem são: “ser”, “aula(s)” e “dia”. Esse texto foi auxiliar à proposta de produção número três: “Voltar de férias é...”. Em função da temática proposta, entendemos que tais palavras também podem ser consideradas tópico dessa proposta.

E, no terceiro texto de apoio, “ser” apareceu novamente, junto com: “escola”, “futuro” e “Brasil”. Nessa proposta de número 5, os alunos foram instigados a escrever – conforme campanha nacional da Rede Globo nos anos de 2018 e 2019, que apresentava em seus telejornais (em forma de vídeo) os desejos dos brasileiros para nosso país – completando a frase: “o Brasil que eu quero para o futuro...”. Portanto, as palavras “futuro” e “Brasil”, para essa proposta, assim como outras nos outros textos de apoio, também podem ser consideradas palavras-tópico. Acreditamos que a palavra “escola” esteja entre as mais frequentes porque os

alunos demonstram a preocupação com a sua escola nos seus textos. Além disso, o redator do texto de apoio também manifesta a melhoria das escolas como um dos seus desejos para o “futuro” do “Brasil”.

Outro item analisado foi a riqueza lexical dos textos dos alunos, que diminuiu à medida que aumentamos o número de textos sob análise, como mostra a Tabela 17:

Tabela 17 – Riqueza Lexical

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
Riqueza Lexical	20,65%	22,12%

Fonte: Elaboração própria.

Como mostra a Tabela 17, a riqueza praticamente não variou entre os textos de uma escola para a outra, pois ficou em torno de 20% e 22%. Essa riqueza diminuiu novamente à medida que aumentamos o número de textos sob análise e diversificamos as temáticas. Mas, ainda assim, é semelhante à riqueza apresentada no estudo piloto II, com os 36 textos sobre a mesma proposta, que era em torno de 29%. Conforme já comentado no segundo estudo piloto, esse percentual de riqueza é semelhante a de um artigo científico da área de biologia, em que geralmente os termos da área são repetidos como forma de reiterar o assunto. Nesse gênero, conforme Finatto, Cremonese e Azeredo (2008), a riqueza fica em torno de 29%.

Os textos de apoio e os textos dos alunos, separados de acordo com a proposta e com o texto de apoio ao qual se relacionam, foram contrastados conforme Figura 6 a seguir:

Figura 6 – Riqueza Lexical – textos de apoio

	Texto de apoio - A vida como filme	Textos da proposta 01 - Escola Municipal	Textos da proposta 01 - Escola Estadual
Riqueza Lexical	52,3%	27,8%	29,9%

	Texto de Apoio – Voltar de férias é...	Textos da proposta 03 – Escola Municipal
Riqueza Lexical	52,3%	13,6%

	Texto de Apoio – O Brasil que eu quero	Textos da proposta 05 – Escola Estadual
Riqueza Lexical	59,3%	15,9%

Fonte: Elaboração própria.

Em todos os grupos de redações dos alunos a riqueza é menor que a dos textos de apoio. É importante lembrar que a medida chamada de riqueza lexical está relacionada com a quantidade de palavras e, conseqüentemente, com o número de textos também. Nesse sentido, é natural que um único texto (como no caso dos textos de apoio) tenha uma riqueza maior que dezoito textos, por exemplo (como no caso da proposta 1). Ainda assim, destaca-se, entre os textos dos alunos, riqueza consideravelmente maior nas redações de proposta 1, mesmo que essa tenha mais quantidade de produções (dezoito) que as propostas 3 e 5, com quatorze e quinze textos, respectivamente. Nesse caso, os resultados foram diferentes do que se imaginava, uma vez que, mesmo com mais textos sob análise, a proposta 1 apresentou maior riqueza lexical do que as outras: 28% em relação a 14% e 16% nas propostas 3 e 5, respectivamente.

Mais uma vez, é interessante observar a diferença de riqueza lexical observada em relação ao texto de apoio “Voltar de férias é...”. Conforme já mencionado, tal texto, na época em que foi escrito, foi redigido por uma aluna, assim como as redações de nosso *corpus* de estudo. Ainda assim, sua variedade vocabular (52,3%) se aproxima dos outros textos de apoio e se diferencia das produções dos alunos (13,6%).

Por fim, foram analisados e contrastados os dados referentes ao *hapax* desse conjunto maior de textos. A amostra dessa camada de palavras de frequência única, com dezessete exemplos entre as primeiras ocorrências das letras A e P, foi sintetizada no Quadro 19:

Quadro 19 – *Hapax* – estudo piloto III

	Escola Municipal	Escola Estadual
Letra A – quantidade	112	140
Letra A – exemplos	abc, aborta, <b>abri, acaba, acabamos</b> , acada, acaso, <b>aceitarem, achando</b> , acidente, <b>acontecer, aconteceria, aconteceriam, acontecido</b> , acordado, acordando, acordava...	aberta, abolir, <b>abre, abril, abrir, abriu</b> , abusada, <b>acabar, aceitam, acharia, achã</b> m, acima, <b>aconteceram, acontecido, acontecimentos, aconteça</b> , acorrentado...
Letra P – quantidade	105	154
Letra P – exemplos	p, padaria, <b>pagar</b> , paisagens, palhaçada, palmares, papo, <b>param</b> , paraná, <b>parecida</b> , parede, participar, participaram, partiu, parto, passam, passamos...	padrão, padrões, paeva, <b>paga, pagamentos, pagando, pagaram, pagariam</b> , palavra, palavras, papa, papeis, <b>parados</b> , parceiro, parceiros, parcelado, <b>parece</b> ...

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

Conforme o Quadro 19, iniciando com a letra A e se repetindo apenas uma vez no *corpus*, observamos 112 palavras num grupo e 140 no outro. Pelos exemplos do Quadro 35 e considerando a base semântica das palavras, cerca de 61% se repetem, que são as palavras grifadas em negrito, todos verbos, conjugados em suas mais diversas formas. Entre esses verbos, alguns têm a mesma base, como, por exemplo: acaba – acabamos, acontecer – aconteceria – aconteceriam – acontecido, abre – abril – abrir – abriu, acharia – achã.

Já iniciando pela letra P e aparecendo apenas uma vez no *corpus*, foram 105 e 154 palavras na escola do município e na escola do estado, respectivamente. A diferença foi de quase 50 palavras a mais na escola estadual, que também foi a que apresentou 1.500 palavras a mais no total. Nos exemplos do Quadro 35, considerando a base semântica, apenas 29,4% das palavras se repetem de um grupo para o outro, como vemos grifadas em negrito. Elas se referem às bases verbais: “pagar”, “parar” e “parecer”, nas suas diversas formas.

Percebemos, quando analisadas as palavras de frequência única, que muitas delas possuíam a mesma base. Portanto, caso a lista fosse lematizada, não

estariam no *hapax*. Em função disso, apresentamos também outro recorte dessa mesma camada do *hapax* das letras A e P, mas, agora, lematizadas, conforme suporte de colegas pesquisadores da USP, para que possamos visualizar as diferenças entre um resultado e outro. Essa lista lematizada, com 20 exemplos em cada, é o que apresenta o Quadro 20:

Quadro 20 – *Hapax* lematizado – estudo piloto III

	Escola Municipal	Escola Estadual
Letra A – quantidade	45	51
Letra A – exemplos	abc, abortar, <b>abrir</b> , acaso, <b>aceitar</b> , acidente, acordado, acre, adaptar, adiantar, adoecer, aeroporto, ah, ajeitar, ajuntar, ajustar, ali, aliás, almoço, aloprar...	<b>aberto</b> , abolir, abusar, <b>aceitar</b> , acima, acontecido, acontecimento, acorrentado, acostumar, admirar, adolescente, adolescência, afeganistão, afundar, agir, agitado, alan, aldeia, alegrar, alar...
Letra P – quantidade	57	58
Letra P – exemplos	p, padaria, <b>pagar</b> , paisagem, palhaçada, palmar, papo, paraná, parede, partir, parto, passe, passeio, passo, paz, país, padrinho, pedro, peito, pele...	pago, <b>pagamento</b> , palavrar, palavra, papa, papar, parcelar, parente, particular, particularmente, peto, peculiaridade, pegado, pensamento, perceber, perigoso, perito, pertencer, pesado, pio...

Legenda: Negrito – palavras semanticamente comuns entre os grupos.

Fonte: Elaboração própria.

No Quadro 20 foi possível visualizar que, ao lematizarmos nosso *corpus*, a camada de *hapax* diminuiu consideravelmente. As quantidades diminuíram, na letra A, de 112 para 45 palavras na escola municipal e de 140 para 51 na escola estadual. Já com relação à letra P, a diminuição foi de 105 para 57 e de 154 para 58. Além disso, como percebemos pelas palavras grifadas em negrito, a coincidência semântica nessa camada também mudou bastante, pois somente dois verbos permanecem como coincidentes na letra A (“abrir” e “aceitar”) e um apenas na letra P (“pagar”). Ou seja, a partir dessa lematização, nossa camada mais promissora, a do *hapax*, ficou mais enxuta. Em função disso, espelhou melhor as especificidades de vocabulário de cada grupo.

Após a lematização, percebemos também que o mesmo grupo de palavras de frequência única, que chegou a ter 64% de coincidência de ocorrências entre os exemplos da letra A, quando não lematizado, diferenciou-se bastante quando

lematizado. Esses 64% de coincidência diminuíram para apenas 10%. Com relação aos exemplos da letra P, a semelhança que era de 29% e passou para apenas 5% (somente o verbo “pagar”). Nesse sentido, nosso resultado tornou-se pontual e mostrou-nos algo diferenciado até então: a diferença entre o vocabulário escrito dos alunos das duas escolas.

Nos estudos piloto anteriores, o vocabulário escrito dos alunos foi semelhante tanto com relação às palavras mais frequentes, quanto com relação ao *hapax*. Além disso, a riqueza lexical também não se distanciava muito de uma escola para outra. No entanto, a partir da lematização, foi a primeira vez que percebemos resultados que distanciam uma escola da outra: *hapax* não coincidentes. Em função disso, no próximo estudo piloto, bem como em nosso experimento principal, só analisaremos *hapax* lematizados.

Em relação ao terceiro estudo piloto, percebemos que, ao aumentar e, principalmente, ao diversificar as temáticas das redações dos alunos, misturando-as, o vocabulário tende a se expandir, tornando mais complicados os contrastes e as comparações. Em função disso, optamos por contrastar pontualmente cada texto de apoio com as redações dos alunos específicas de cada proposta. Nesse sentido, a comparação foi mais relevante, pois resultou em observações representativas.

Além disso, permanecemos com a ideia de aprofundar o estudo da camada *hapax legomena*. Neste estudo piloto III, percebemos ainda mais a necessidade da lematização dos itens, uma vez que muitas das palavras que se apresentaram no *hapax* não estariam ali caso nossas contagens fossem lematizadas. Então, lematizamos nossas listas, obtendo *hapax* com menos quantidade de palavras, pois, assim, a camada espelhou de fato as especificidades de cada grupo, mostrando que praticamente não há coincidência no que se refere à individualidade de cada grupo.

Por fim, pretendemos inserir textos do 6º ano/B30 na pesquisa, como distrator, no sentido de avaliarmos o que e o quanto permanece ou muda à medida em que esses alunos, das mesmas redes e escolas, vão crescendo, amadurecendo e avançando na vida escolar. Esses textos do 6º ano/B30 são todos sobre a mesma proposta de produção textual e serão apresentados a seguir.

### 5.3.4 Estudo Piloto IV

Este estudo piloto pode ser considerado uma derivação temática do experimento principal, que se ocupa de verificar semelhanças e diferenças entre o vocabulário escrito de dois grupos: os estudantes da escola municipal e os estudantes da escola estadual. Dado que, em meio a amostra de estudo, há textos produzidos por alunos que estão em anos escolares diferentes, aproveitamos para verificar em que medida o tempo de escolaridade maior ou menor repercutiria sobre o perfil do vocabulário escrito dos estudantes.

Durante a organização e categorização do *corpus* de estudo, constatamos que, em meio ao acervo de redações dos alunos do 8º ano/C20, havia também produções dos alunos de 6º ano/B30. Esses textos do sexto ano totalizavam 60 produções.

Em busca de uma possível diferenciação do vocabulário empregado de acordo com o ano/ciclo escolar, nosso estudo piloto IV realizou mais um contraste segmentado. Quisemos mensurar o quanto do vocabulário desses alunos permanece ou se modifica à medida em que eles crescem, amadurecem e avançam na vida escolar. Em função disso, comparamos o vocabulário escrito das 60 redações do 6º ano/B30 com o das 94 redações do 8º ano/C20. Nosso *corpus* do quarto estudo piloto ficou, então, organizado da seguinte forma:

Quadro 21 – Corpus do estudo piloto IV

<b>Redações do 6º ano/B30</b>	<b>Redações do 8º ano/C20</b>
60 textos	94 textos

Fonte: Elaboração própria.

Como as quantidades de redações dos alunos dos dois grupos eram diferentes, foi preciso ponderar as comparações. Então, foi necessário aplicar uma razão de proporção entre os textos envolvidos. Desse modo, pudemos perceber o quanto as redações do grupo do 8º ano (em maior quantidade) representam em relação aos textos do grupo do 6º ano (em menor quantidade). Essa medida de proporção ficou assim:

$\frac{94 \text{ textos do } 8^\circ \text{ ano}}{60 \text{ textos do } 6^\circ \text{ ano}} = 1,56$
--

Ou seja, quando multiplicamos os 60 textos do 6º ano por 1,56, temos o total de 94 (quantidade de textos do outro grupo). Isso quer dizer que as redações do 8º ano representam 1,56 das redações do 6º ano. Em outras palavras, em porcentagem, os textos do 8º ano representam 56% a mais do que os textos do 6º ano.

Dessa forma, a fim de equilibrar os dados dos dois anos escolares, nesse estudo piloto IV, nossos resultados brutos, como número de *tokens* e de *types*, quando relacionados aos 60 textos dos alunos do 6º ano, foram todos multiplicados por 1,56 para que fosse possível a proporção ou uma escala de ambos os anos no mesmo patamar. É importante ressaltar que todos esses cálculos foram feitos com objetivo de estimar sempre a média ponderada dos resultados obtidos por meio das redações dos alunos.

Com isso, comparamos a quantidade total de palavras e a quantidade de palavras diferentes por ano/ciclo, apresentados na Tabela 18:

Tabela 18 – *Tokens/types* – em ponderação – estudo piloto IV.

	<b>6º ano/B30</b>	<b>8º ano/C20</b>
<i>Tokens</i>	12.662	18.307
<i>Types</i>	2.634	3.188

Fonte: Elaboração própria.

Conforme visualizamos no Quadro 38, os alunos do 8º ano/C20 escrevem, de acordo com a nossa ponderação, mais palavras. Essa diferença é expressiva, pois chega-se a 5.645 *tokens* a mais nas redações do 8º ano.

Em relação ao número de palavras diferentes, os alunos do 8º ano seguem escrevendo mais, pois apresentaram 554 *types* a mais que os alunos do 6º ano. Essa diferença foi menor do que quando considerados os *tokens*, mas, ainda assim, foi considerável.

Acreditamos que seja natural que educandos mais velhos e que já avançaram em seus estudos escrevam mais em suas redações. Ainda assim, essa comparação é válida, pois nos mostra o quanto mais eles escrevem e em quais itens exatamente estão essas diferenças.

A riqueza lexical em contraste nos dois anos/ciclos está resumida na Tabela 19, a seguir:



Tabela 19 – Riqueza Lexical – estudo piloto IV

	6º ano/B30	8º ano/C20
Riqueza Lexical	20,80%	17,41%

Fonte: Elaboração própria.

Conforme já mencionado, com uma diferença de dois anos de vida estudantil, os alunos tendem a escrever mais e a usar mais palavras diferentes. Isso ficou evidente, inclusive, nos dados da Tabela 18. Entretanto, conforme a Tabela 19, em termos proporcionais, quando relacionamos essas quantidades, a diversidade do vocabulário ficou somente um pouco menor, em torno de 3%, o que não pode ser considerada diferença considerável. Ou seja, embora os estudantes escrevam mais ao longo dos anos escolares, a variedade vocabular exibida não parece ser expressivamente diferente. Isso aponta para uma variação menor que a de um artigo científico na área de biologia, que usamos como referência e exibe cerca de 29% de riqueza lexical (FINATTO; CREMONESE; AZEREDO, 2008).

Para compararmos os graus de frequência e a proporção de palavras gramaticais e de palavras lexicais dos diferentes anos/ciclos, foi preciso colocar essas palavras mais frequentes lado a lado. O Quadro 22 mostra as 50 palavras mais frequentes, divididas entre gramaticais e lexicais, conforme a ordem em que aparecem:

Quadro 22 – 50 palavras mais frequentes 6º/8º ano

	6º ano/B30	8º ano/C20
Palavras Gramaticais (amostra)	<b>o, que, e, eu, não, mais, de, um, no, para, do, com, sem, ele, em, por, se, muito, isso, todo,</b> porque, <b>bom, pro,</b> melhor, menos, também, esse, <b>ou, mas,</b> hoje	<b>o, e, eu, que, um, ele, de, do, no, não, para, com, mais, muito, se, em, por, todo, mas,</b> depois, quando, <b>bom, ou, pro,</b> tudo, outro, como, até, pelo, então
Palavras Lexicais (amostra)	<b>ter, brasil, querer, ser, pessoa, escola,</b> futuro, rua, <b>ir,</b> lixo, <b>estar,</b> poder, <b>coisa,</b> professor, parar, <b>fazer, casa, dia, ficar,</b> podar	<b>ser, ter, ir,</b> filme, <b>pessoa,</b> amigo, <b>fazer, casa,</b> vida, <b>estar, coisa, dia, ficar, querer,</b> pai, <b>escola,</b> achar, <b>brasil,</b> ano, voltar

Legenda: Negrito – palavras semanticamente comuns entre os dois grupos.

Fonte: Elaboração própria.

Por meio do Quadro 22, observamos que, entre as 50 mais frequentes, a proporção de palavras gramaticais em relação à de lexicais é a mesma nos dois anos, pois ambos apresentaram 30 palavras gramaticais e 20 palavras lexicais. Já

com relação às palavras coincidentes semanticamente nos dois grupos, grifadas em negrito, a porcentagem foi relativamente alta, pois 73,3% das palavras gramaticais usadas pelos dois anos/ciclos foi a mesma. Isso mostra que os recursos gramaticais utilizados com preferência por esses redatores não tendem a se modificar ao longo desses dois anos de vida estudantil.

Além disso, 65% das palavras lexicais utilizadas são comuns entre os dois grupos. Isso demonstra aproximação no vocabulário escrito mais usado pelos estudantes, independente do ano/ciclo e da temática das produções (que eram variadas). Percebemos que muitas dessas palavras lexicais são verbos, dentre os mais usados e produtivos na língua (BIDERMAN, 1998a, p. 12), como “ser”, “ter”, “ir”, também usados, por vezes, como verbos auxiliares. Conforme Biderman (1998a, p. 14), “esses [...] verbos registram altíssima frequência por serem reiterados continuamente no texto. Dada a centralidade do verbo na articulação do discurso, é normal a enorme repetição dessas palavras”. Além desse aspecto, acreditamos ser importante destacar a coincidência, entre os dois grupos, dos substantivos: “casa”, “dia”, “brasil”, “escola”, “pessoa” e “coisa”. Esse repertório lexical coincidente aponta semelhança nos assuntos de preferência dos alunos. Isso acontece, como vimos em estudos piloto anteriores, nos grupos das duas escolas e, como vemos neste estudo piloto, independente do ano/ciclo ao qual os redatores pertencem.

Outro elemento em destaque na distribuição de frequências foi o grupo de itens que compõem o *hapax*. Nosso recorte de observação incidiu sobre as palavras de única ocorrência iniciadas pela vogal A e pela consoante P, como mostra o Quadro 23:

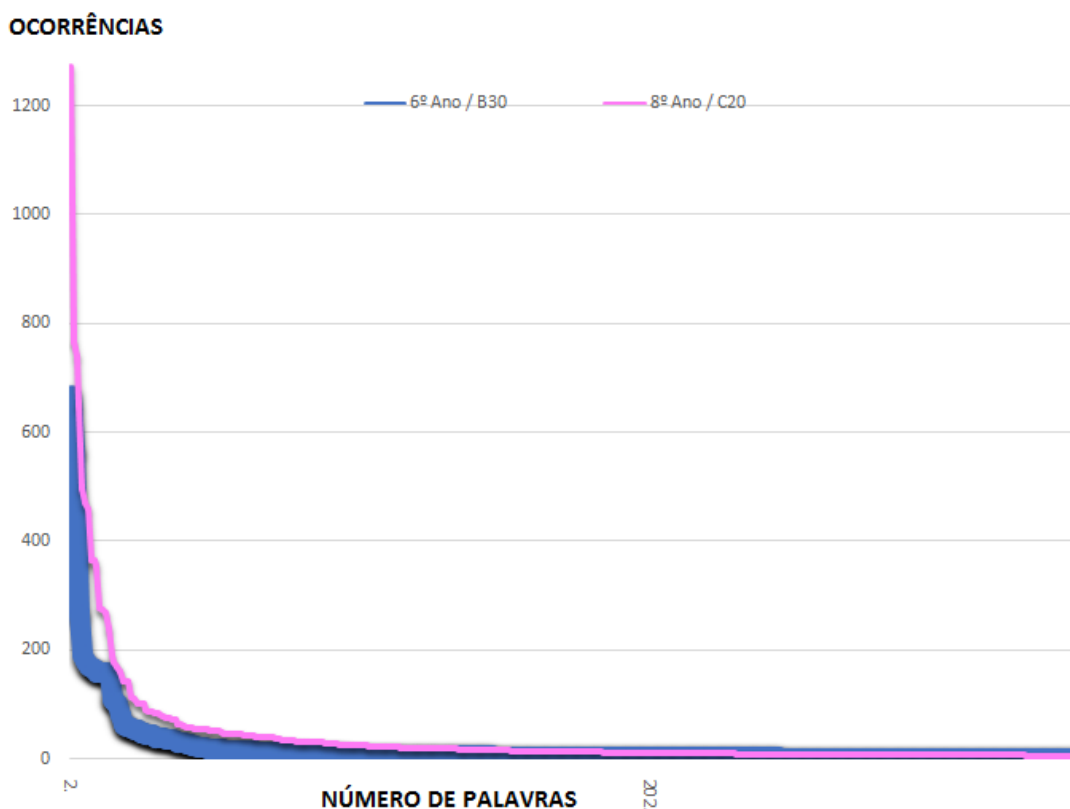


acontecido/acontecimento, passe/passo, pedro/pedrinho, entre outros. Além disso, percebemos também ocorrências como: *apoucar*, *atendar*, *ponhar*, *potenciar*, entre outros. Tais formas foram automaticamente classificadas pelo lematizador como as formas canônicas de “atenda”, “apouco”, “ponha” e “potência”, pelo que observamos em buscas no *corpus*. Não quisemos corrigi-los manualmente. Essa ferramenta, como é natural, tem uma margem de erro de reconhecimento de lemas/flexões. Entretanto, seu percentual de erros não supera o de acertos. Assim, a margem de erros “embutidos” ao sistema pode afetar os resultados, mas acreditamos que sua eficiência justifica essa incorporação de “defeitos”.

Por outro lado, no recorte de *hápax* iniciados por P, tivemos maior coincidência entre os usos dos dois grupos, cerca de 17,4% das palavras foram repetidas, independente do ano/ciclo dos alunos. Ainda que essa porcentagem tenha sido maior, em relação ao resultado entre os *hapax* da letra A, a coincidência permanece pequena entre os grupos quando se trata de especificidades do seu repertório vocabular escrito.

Para que fosse possível visualizar melhor o perfil do vocabulário escrito empregado pelos estudantes dos dois anos/ciclos, o Gráfico 3 abaixo mostra o gráfico de frequência das palavras utilizadas, bem como os itens da camada *hapax legomena*:

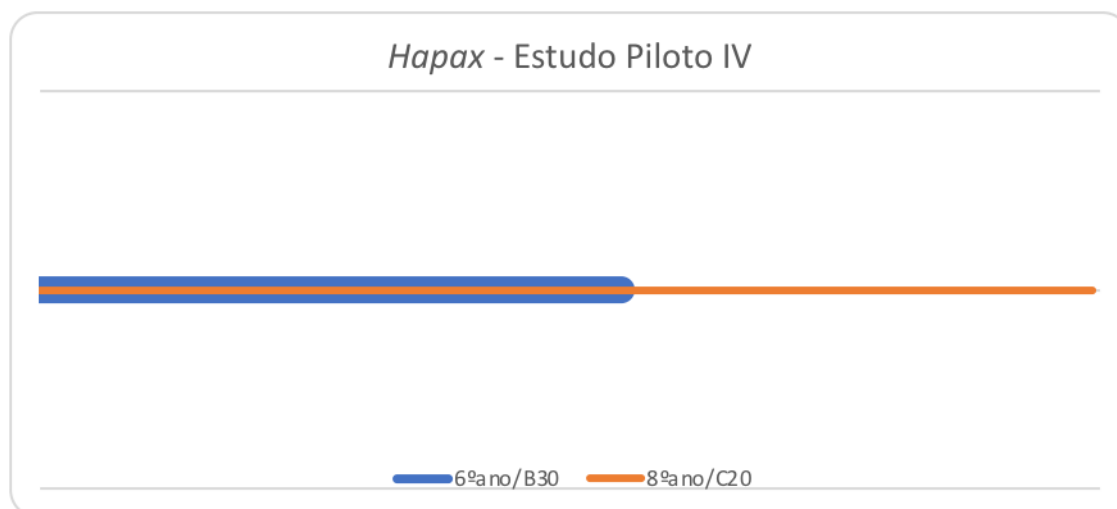
Gráfico 3 – Gráfico de frequência de palavras – Estudo Piloto IV



Fonte: Elaboração própria.

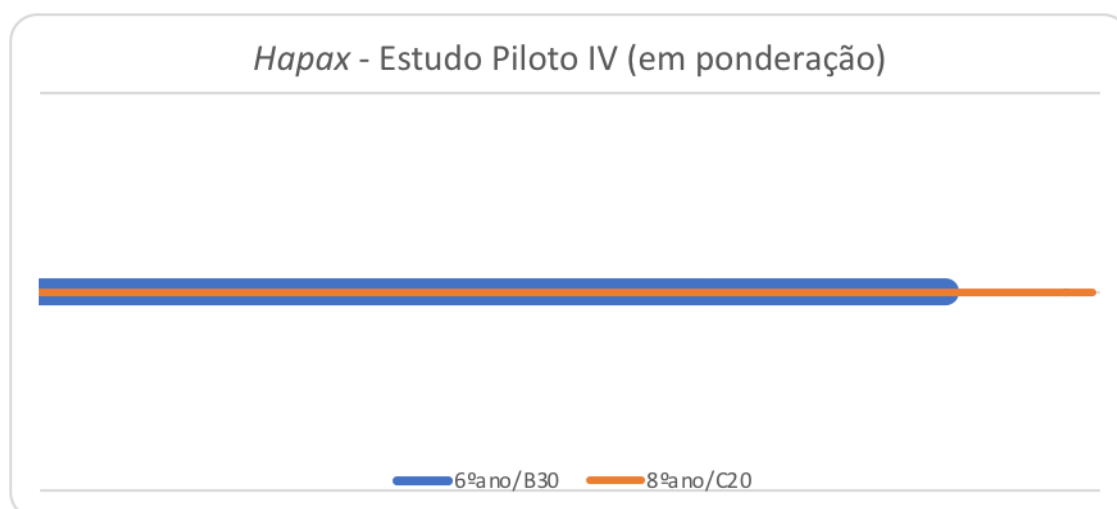
A partir do Gráfico 3, é possível perceber que o início de cada traço remete à palavra gramatical mais frequente em ambos os grupos: “o” – que no 8º ano/C20 tem 1.271 ocorrências, e no 6º ano/B30 tem 699 ocorrências. Mesmo com essa diferença, natural em função do *corpus* do 8º ano/C20 ter maior quantidade de palavras, notamos que as curvas dos dois grupos não diferenciam muito uma da outra, ou seja, seus dados, em proporção, realmente se aproximam, por isso seus gráficos também são semelhantes. Além disso, a camada mais extensa, independente do ano/ciclo, é a das palavras de frequência única, que se expande para além do Gráfico 3, anteriormente demonstrado.

Para que fosse possível melhor visualização dessa camada *hapax*, a seguir, o Gráfico 4 mostra o recorte dessa camada nos dois grupos:

Gráfico 4 – *Hapax* – Estudo Piloto IV

Fonte: Elaboração própria.

Como é possível perceber por meio do Gráfico 4, o *hapax* do 6º ano/B30 tem praticamente a metade da quantidade de palavras do *hapax* do 8º ano/C20. Portanto, a camada de palavras de frequência única praticamente dobrou em dois anos de estudo. Entretanto, não podemos esquecer que esse resultado precisa ser relativizado na medida em que compara número maior de textos (e, conseqüentemente, de palavras) do 8º ano do que do 6º ano. Ao ponderarmos as comparações da camada *hapax*, assim como fizemos com *tokens* e *types*, o gráfico passa a ser o seguinte:

Gráfico 5 – *Hapax* – Estudo Piloto IV (em ponderação)

Fonte: Elaboração própria.

Conforme é possível visualizar no Gráfico 5, ao ponderarmos os números relativos ao *hapax* dos textos analisados, a diferença entre os dois grupos diminui consideravelmente. Isso acontece porque multiplicamos a quantidade de *hapax* do 6º ano/B30 por 1,56, para que seja possível uma equivalência entre os dois grupos, que têm quantidades de textos diferentes. Ao multiplicarmos, o *hapax* desse grupo, que era de 420 itens, passa para 655 itens. Dessa forma, percebemos que a diferença entre a quantidade de especificidades entre os dois anos/ciclos, quando ponderadas, não são tão relevantes.

A partir do quarto e último estudo piloto, então, concluímos que os estudantes escrevem mais à medida que avançam na vida escolar, do 6º ano para o 8º ano. Além disso, também escrevem palavras diferentes em maior número. Entretanto, quando relacionados apenas os números brutos, sem ponderação e sem observação específica por tipos de frequências, a conclusão seria a de que não houve variação expressiva. Os escores de riqueza lexical de 20% e 17% aproximam os dois grupos, independentemente das séries/anos. Essa diferença parece ser mínima. Ao observarmos os itens *hapax*, conforme o Gráfico 5 acima, vemos que o repertório do vocabulário não repetido praticamente dobra de extensão em dois anos de escolaridade, entretanto, quando ponderamos os dados, novamente os resultados voltam a se aproximar.

Desse modo, a lematização dos itens e o olhar atento aos *hapax* foram procedimentos que muito contribuíram para esse último estudo parcial, acurando nossos resultados. Assim, partimos da expectativa de que isso seria relevante também no experimento principal – com todas as redações, temáticas, ano/ciclos e quantidades de textos disponíveis no *corpus* total de estudo.

#### 5.4 PARTE II: EXPERIMENTO PRINCIPAL

Após o desenvolvimento dos estudos piloto, realizamos o experimento com todas as redações do *corpus*. Os estudos piloto serviram para repensar formas e métodos de análise, ferramentas computacionais utilizadas, contrastes realizados, quantidade e temática dos textos, entre outras questões.

Conforme já referido, nosso *corpus* tem 154 redações de alunos de 6º e de 8º ano de duas escolas públicas do RS: uma municipal e outra estadual. O CorPop

(PASQUALINI, 2018) foi selecionado como *corpus* de referência, conforme já explicado anteriormente. O Quadro 24, a seguir, sintetiza os dados do *corpus* de estudo e do de referência.

Quadro 24 – Composição geral do *corpus* – experimento principal

<b>Corpus de estudo</b>		<b>Corpus de referência</b>
Escola Municipal	Escola Estadual	CorPop
77 redações	77 redações	<i>Corpus</i> de referência do português popular brasileiro escrito

Fonte: Elaboração própria.

A seguir, apresentamos uma descrição do *corpus* de estudo, em diferentes contrastes, por diferentes dimensões e por tipos de formas ou camadas de frequência do vocabulário.

#### a) Número total de palavras e palavras diferentes

O primeiro resultado obtido foi o número total de palavras e o número de palavras diferentes. Isso é o que vemos na Tabela 20:

Tabela 20 – Relação *token/types* – experimento principal

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
<i>Tokens</i>	14.735 (média de 191/texto)	15.732 (média de 204/texto)
<i>Types</i>	2.444 (média de 96,5/texto)	2.609 (média de 101,4/texto)

Fonte: Elaboração própria.

Na Tabela 20, acima, percebemos que de fato os alunos da escola estadual escrevem mais do que os alunos da escola municipal, quando considerado o total de textos da nossa pesquisa. Esse fato só não ocorreu no estudo piloto I, quando analisamos somente quatro redações de cada escola. Em todos os outros estudos, incluindo este experimento principal, com todo nosso acervo, o número de *tokens* foi sempre maior na escola estadual. Observamos cerca de 1.000 palavras a mais nos 77 textos da escola estadual, uma diferença de aproximadamente 13 palavras por texto a mais quando considerada a média de palavras por texto.

A diferença entre o número de *types* foi menor, pois os textos da escola municipal apresentaram 165 *types* a menos do que os da escola estadual. A média



de *types* por texto não teve diferença considerável, pois variou de 96,5 na escola municipal para 101,4 na escola estadual, apenas.

### b) Proporção de palavras lexicais e gramaticais

Em relação aos graus de frequência e à proporção de palavras lexicais e de palavras gramaticais, o Quadro 25 expõe os dados das 50 primeiras palavras mais frequentes escritas pelos dois grupos:

Quadro 25 – Proporção e frequência /palavras gramaticais e lexicais – experimento principal

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
Palavras Gramaticais	<b>o, que, e, eu, de, a, as, para, mais, um, os, com, no, se, na, uma, por, minha, em, do, ou, ele, da, sem, porque, muito, pra, todos, quando, tudo, meu, bom, mas, me, depois, das, ela, isso</b>	<b>o, que, e, a, de, um, eu, para, os, as, mais, com, eles, do, em, uma, por, no, se, da, mas, muito, na, minha, sem, meu, ele, isso, como, depois, quando, todos, sua, seu, bom</b>
Palavras Lexicais	<b>pessoas, brasil, filme, quero, ter, dia, ser, vida, escola, coisas, casa, voltar</b>	<b>brasil, pessoas, quero, filme, tem, casa, futuro, come, seria, queria, tesouro, foi, dia, escolas, ser</b>

Legenda: Negrito – palavras semanticamente comuns entre os dois grupos.

Fonte: Elaboração própria.

Conforme o Quadro 25, percebemos que a proporção entre palavras gramaticais e palavras lexicais é semelhante nos dois grupos de textos, uma vez que no grupo da escola municipal, 38 palavras foram gramaticais, enquanto 12 foram lexicais. Por sua vez, na escola estadual, 35 são gramaticais e 15 são lexicais, números próximos nos dois grupos. Além disso, grifadas em negrito, as palavras coincidentes foram cerca de 85% entre as gramaticais, o que comprova que os recursos gramaticais utilizados por esses alunos são muito próximos nos dois grupos. Nenhum deles utiliza recursos coesivos diferenciados como “todavia” ou “conquanto”, por exemplo, com frequência.

Já em relação às palavras lexicais, a semelhança foi em torno de 74%. Muitas dessas coincidências lexicais seguem sendo verbos, alguns inclusive usados como verbos auxiliares. Entre os substantivos recorrentes nos dois grupos, estão: “pessoa”, “Brasil”, “casa”, “dia” e “escola”. Tais substantivos demonstram que os assuntos tratados nas redações de ambos os grupos são semelhantes, ainda que as propostas de produção tenham sido variadas. Isso já havia sido verificado nos estudos piloto anteriores.

### c) Riqueza lexical

Outro item analisado foi a riqueza lexical dos textos das duas escolas. Essa medida estatística que relaciona a quantidade de palavras diferentes com a quantidade total de palavras de um *corpus* pode ser conferida na Tabela 21, abaixo:

Tabela 21 – Riqueza Lexical – experimento principal

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
Riqueza Lexical	16,5%	16,5%

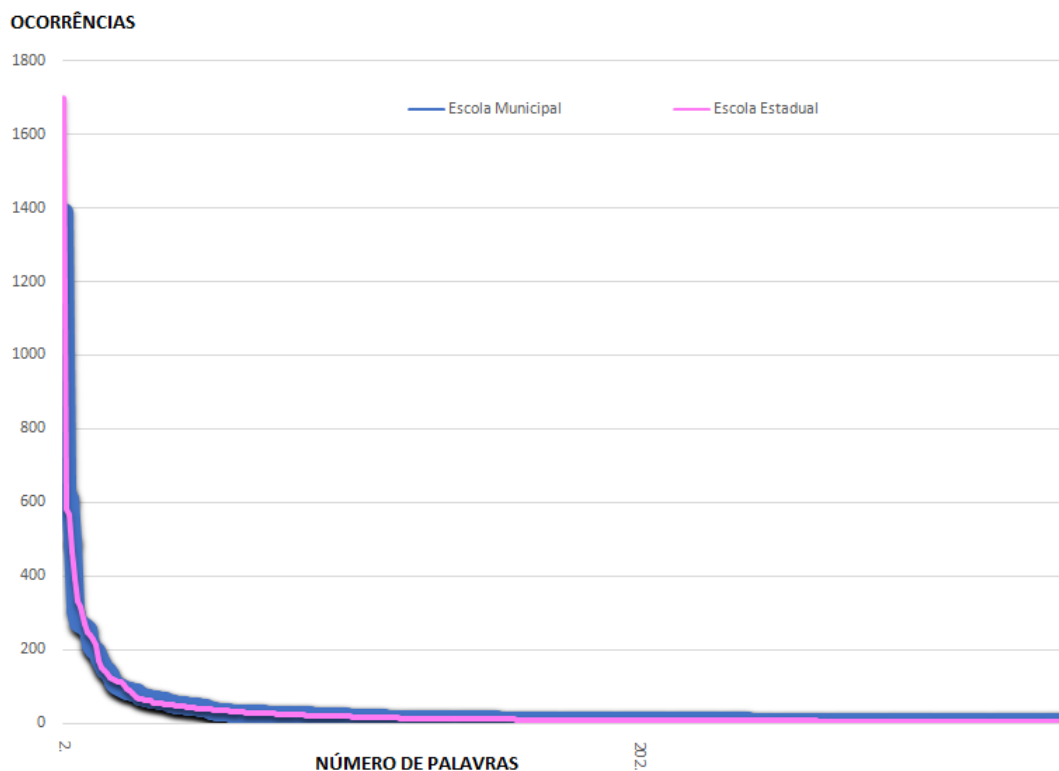
Fonte: Elaboração própria.

A riqueza lexical encontrada foi exatamente a mesma nas duas escolas. Conforme fomos aumentando o número de textos do nosso *corpus* de estudo, as riquezas foram diminuindo, pois, como já mencionado, é natural que o maior número de textos tenha vocabulário mais limitado. Nesse caso, a riqueza dos 77 textos, tanto da escola municipal, quanto da escola estadual, foi de iguais 16,5%. Essa medida aponta que os alunos são bastante repetitivos ao escreverem seus textos, não variando muito o seu vocabulário escrito. Conforme Evers (2013, p. 129), “[...] quanto mais Básico é o nível do conjunto de textos, mais palavras distintas são usadas”. Nesse sentido, muitos dos estudos com estatística vocabular apontam para o fato de que um texto considerado mais proficiente (no caso da proficiência do Celpe-Bras, em Evers, 2013) ou com melhor avaliação e maior nota (no caso das redações do vestibular da UFRGS-2014, em Evers, 2018) não é aquele com vocabulário mais variado, conforme poderíamos pensar. Contrariando as expectativas, o que os estudos mostram é que geralmente quem muito diversifica as palavras em seu texto não as emprega corretamente ou não as conhece bem. Ainda, segundo Evers (2013, p. 129), textos com nível mais avançado repetem as palavras e obtêm melhores notas, “talvez porque, justamente por se repetirem, acabam construindo um texto mais coeso, e, portanto, mais claro, resultando em um desempenho geral melhor [...]”.

Os resultados de nossa variedade vocabular, portanto, demonstram que os textos dos alunos do Ensino Fundamental de escolas públicas analisados aproximam-se, nesse quesito, dos textos com proficiência avançada do exame Celpe-Bras-2006 e das redações do CV-UFRGS-2014 com melhores notas. E, segundo nossos dados estatísticos, isso acontece igualmente nas duas escolas.

Para demonstrar as frequências, utilizamos novamente o gráfico de frequência de palavras. Observemos o Gráfico 6, a seguir:

Gráfico 6 – Gráfico de frequência de palavras – Experimento Principal

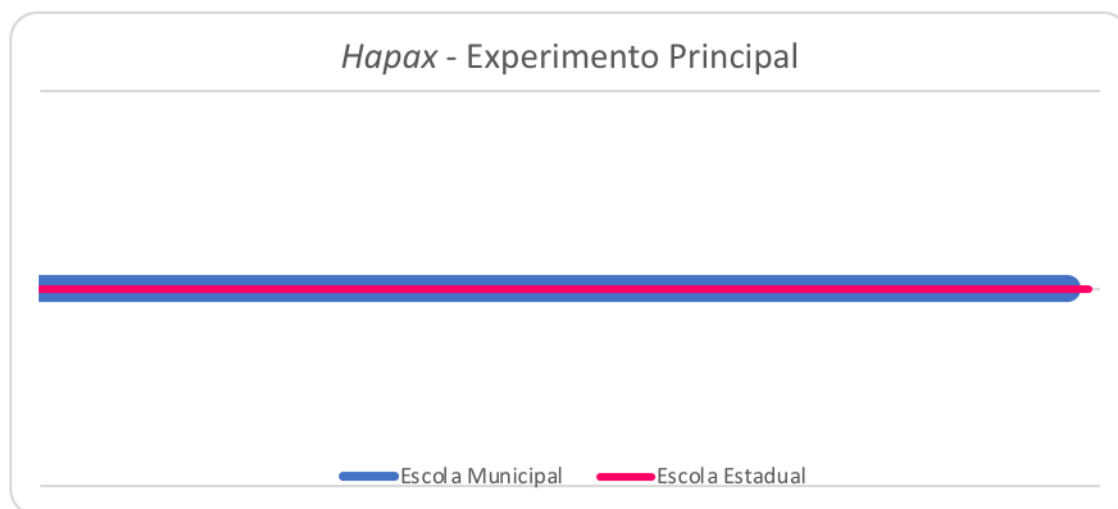


Fonte: Elaboração própria.

Esse gráfico mostra tanto a quantidade total de palavras, quanto o que chamamos de camadas, como já explicamos anteriormente. Então, é possível visualizar, agora sob a forma de gráfico, que os resultados não variam muito de um grupo de textos para o outro.

#### d) *Hapax*

A partir do gráfico, é possível perceber que os textos escritos por alunos da escola estadual têm de fato maior número de palavras, mas que ambas as frequências e variedades são muito próximas, conforme suas curvaturas. A cauda, que representa o grupo das palavras de frequência única, é a parte mais extensa em ambos os grupos, também continuando para além da imagem mostrada no Gráfico 6, como no gráfico do *hapax* do estudo piloto IV. Em função disso, o Gráfico 7, a seguir, mostra somente essa parte dos *hapax* dos dois grupos:

Gráfico 7 – *Hapax* – Experimento Principal

Fonte: Elaboração própria.

A partir do Gráfico 7, é possível perceber que a quantidade de palavras de frequência única praticamente não varia entre os dois grupos. Na escola municipal temos 530 palavras com única ocorrência e, na escola estadual, 541. Em mais esse quesito, então, os grupos se aproximam.

Quanto às palavras que aparecem nessa camada *hapax*, temos diferença entre os grupos. Nossa lista lematizada de palavras de frequência única, das letras A e P, foram dispostas no Quadro 26:

Quadro 26 – Hapax legomena – experimento principal

	<b>Escola Municipal</b>	<b>Escola Estadual</b>
Letra A – quantidade	54	57
Letra A – exemplos	abaixar, abandonado, abandonar, abc, abortar, abuso, acaso, <b>aceitar</b> , acender, acidente, acordado, acre, acumulado, acumular, adaptar, adoecer, adubar, aeroporto, afeto, afinal, <b>agradecer</b> , agir, ajeitar, ajuntar, ajustar, alagado, alberto, alicerce, aloprar, altura, amarelo, amazonas, amizade, angustiado, anoitecer, antigamente, apoucar, <b>apresentar</b> , aprontar, areia, ariar, arrecadar, arriscar, assalto, assar, <b>assustado</b> , <b>assustador</b> , atender, <b>atirado</b> , atitude, atravessar, atualmente, atuar, azar	aberto, abolir, abrigo, <b>aceitar</b> , acima, acontecido, acontecimento, acorrentado, acostumar, adiantar, admirar, adorar, afundar, agir, agitado, agradar, <b>agradecer</b> , alan, aldeia, alegrar, alegria, alimento, almoçar, amadurecimento, animado, ante, aonde, apaixonado, apaixonar, aparelho, apegar, apelido, apesar, apoio, aposentadoria, <b>apresentar</b> , aproveitar, aqui, arriscar, artista, artur, aspecto, aspirante, assassinato, assombrado, <b>assustar</b> , asterisco, atear, atencioso, atendimento, <b>atirar</b> , ativar, atrapalhar, atrasado, atrasar, atual, auto
Letra P – quantidade	52	47
Letra P – exemplos	<b>padaria</b> , paisagem, palha, palmar, papo, partir, parto, passe, passeio, passo, pedofilia, pedrinho, pedro, peito, pele, pensamento, pensativo, perdido, pergunta, periferia, <b>perigo</b> , <b>perigoso</b> , permitir, perna, persistir, piada, picape, pintar, pior, <b>piscina</b> , pisar, <b>play</b> , policial, ponhar, pôr, porra, porrada, postar, poupar, <b>povo</b> , pracinha, precioso, prejudicar, <b>presidente</b> , preto, primeiramente, privado, progresso, proibir, prometer, protagonista, provavelmente	<b>padaria</b> , pago, pagamento, palavrar, papa, papar, par, parcelar, parede, parente, particularmente, pato, peculiaridade, pegado, <b>perigoso</b> , perito, pertencer, pesado, piano, <b>piscina</b> , pista, plano, <b>play</b> , ponta, posar, possuir, potenciar, poupar, <b>povoado</b> , prancha, praticamente, praticar, preferido, preocupar, presente, <b>presidir</b> , produto, profissional, profundo, pronto, proposto, proteger, protesto, prova, provavelmente, psicologia, polir

Legenda: Negrito – palavras semanticamente comuns entre os dois grupos.

Fonte: Elaboração própria.

Como podemos ver no Quadro 26, das 111 palavras dos *hapax* dos dois grupos, iniciadas pela vogal A, apenas 11 coincidem semanticamente. Então, somente 9% do *hapax* dessa vogal é próximo nos dois grupos, fato que também aconteceu com a consoante P. Com essa letra, das 99 palavras iniciadas por ela e que ocorreram somente uma vez no *corpus*, 12 palavras têm coincidência semântica. Ou seja, apenas 12%.

Ainda sobre a camada de itens de frequência única, além de ser aquela que nos mostrou a maior diferença entre os dois principais grupos, também é aquela que mostra a extensão do vocabulário desses alunos. Nesse sentido, acreditamos que analisar esses itens foi importante para o nosso trabalho, na medida em que acreditamos que quanto maior for o repertório vocabular de uma pessoa, com mais facilidade ela tende a compreender novas palavras. Então, uma diversidade de itens

nesse grupo *hapax* permite que o indivíduo compreenda novas formas por meio do contexto e de outras palavras conhecidas, por exemplo. Isso vai ao encontro do objetivo das nossas atividades de ensino, que serão sugeridas no Capítulo 7, e que visam ao alargamento do repertório vocabular desses alunos como forma de melhorar também as capacidades de leitura, de escrita e de interpretação desses estudantes.

A partir dos últimos resultados, expostos no Quadro 26, percebemos, então, que as coincidências apontadas entre os dois grupos com relação ao número de *tokens* e de *types*, à riqueza lexical e até mesmo às palavras mais frequentes (e proporção entre gramaticais e lexicais), não se concretizou com relação aos itens do *hapax*. Em relação aos usos específicos de cada grupo, nos recortes das letras A e P, os alunos das duas redes educacionais têm preferências diferentes de vocabulário escrito, o que acaba diferenciando esses perfis de vocabulário somente nesse quesito.

### e) Contraste com CorPop

Neste experimento principal, além do contraste entre as escolas, optamos por confrontar nosso *corpus* de estudo com o CorPop, conforme já mencionado. Essa comparação teve por objetivo verificar em que medida o vocabulário dos alunos se aproxima (ou não) do vocabulário popular escrito do português do Brasil, representado pelo CorPop. Então, as listas lematizadas de ambos os *corpora* foram dispostas lado a lado e, por meio de planilhas do programa *Microsoft Office Excel*, obtivemos quantas e quais palavras eram semelhantes. Isso é o que podemos ver no Quadro 27, abaixo, com amostra das 50 primeiras palavras semanticamente comuns nos dois *corpora*:

Quadro 27 – Palavras semanticamente comuns entre o *corpus* de estudo e o CorPop

<p>abaixar – abandonar – abandono – aberto – abraçar – abrigo – abrir – absurdo – abusar – abuso – acaba – acaso – aceitar – acender – achar – acidente – acima – acontecer – acontecimento – acordar – acordo – acostumar – acreditar – acumular – adaptar – adiantar – admirar – adoecer – adolescente – adorar – adotar – adulto – aeroporto – afeto – afinal agente – agir – agitar – agora – agradar – agradecer ainda – ajeitar – ajudar – ajustar – aldeia – alegrar – alegre – alegria – algo</p>
---

Fonte: Elaboração própria.

A lista lematizada do CorPop que utilizamos nesta comparação também foi gerada automaticamente e já havia sido previamente filtrada por Pasqualini (2018). Em um processo de “limpeza”, foram retiradas dela algumas palavras como, por exemplo, preposições, artigos, alguns pronomes, advérbios, entre outros. Pasqualini, (2018, p. 87), explica que:

A retirada de elementos gramaticais deu-se em virtude de que tais palavras costumam ser as mais comuns em qualquer lista de frequências, pois são os elementos que dão forma aos enunciados e, por isso, se repetem bastante. O mesmo pode ser dito sobre a retirada dos dias da semana e dos nomes de cidades, países e localidades, pois não acrescentam muito ao repertório lexical que pretendemos criar [...].

Em função disso e da margem de erro embutida na lematização do CorPop, ao contrastarmos os dois *corpora*, percebemos que as palavras gramaticais, na sua maioria, não aparecem entre as coincidências dos dois grupos, como é possível perceber a partir do Quadro 25. As palavras encontradas nos dois grupos foram, na sua maioria, palavras lexicais, em função dessa limpeza do CorPop. Ainda assim, acreditamos que este contraste foi produtivo, pois mapeamos as palavras de conteúdo lexical presentes nos dois grupos, obtendo mais um resultado para nosso estudo.

Ao compararmos as duas listas lematizadas, concluímos que 34,1% dos *corpora* contrastados é igual, ou seja, 2.392 palavras se repetem nos dois. Já quando focamos na realidade do nosso *corpus* de estudo, essa porcentagem aumenta para 67,4%. Ou seja, 1.196 palavras, mais que a metade das palavras do nosso *corpus* de estudo, estão também presentes no *corpus* do português popular escrito do Brasil. Isso aponta para o fato de que o vocabulário escrito dos alunos tende a ser bastante próximo do CorPop e tem a maioria das suas palavras semanticamente comuns com o principal *corpus* de referência adotado nesta pesquisa.

## 5.5 KEYWORDS: AS PALAVRAS-CHAVE

O exame das *keywords* serviu para que pudéssemos dimensionar a especificidade do *corpus* de estudo frente ao todo da língua portuguesa escrita.

Nosso *corpus* de estudo, com 26.424 *tokens*, foi contrastado com três *corpora*: Banco do Português, Lácio Web e Corpus Brasileiro.

É importante lembrar que as palavras-chave não necessariamente são palavras que ocorrem somente em nosso *corpus* de estudo. De acordo com Viana (2010, p. 62):

É fundamental ressaltar que as palavras-chave não são formas que ocorrem exclusivamente em um *corpus* ou em outro [...]. Tanto na teoria como na prática, elas são passíveis de serem encontradas em ambas as coletâneas de textos. A questão que se coloca aqui é que as palavras-chave são empregadas com maior frequência relativa em um *corpus* do que em outro.

Dessa forma, interessou-nos observar as palavras-chave positivas, uma vez que “[...] são aquelas que servem para identificar linguisticamente o *corpus* que se deseja descrever. Os resultados incluídos nessa listagem são os que ocorrem com maior frequência comparativa no *corpus* de estudo” (VIANA, 2010, p. 62). Dentre essas palavras estão aquelas que ocorrem em nosso *corpus* de estudo, com frequência distintiva da frequência no *corpus* de referência.

A chavicidade é, assim, um fator a mais para caracterizar o universo vocabular sob exame. Conforme Kader e Richter (2013, p. 20):

As palavras-chave são úteis na análise linguística para fins diversos, tais como:

- identificar a temática de um *corpus* ou de um texto;
- descrever a organização interna de textos;
- localizar marcas indicativas de posicionamento ideológico;
- traçar um perfil lexical de um autor ou de outros indivíduos.

Dessa forma, utilizamos também as listas de palavras-chave obtidas por meio dos nossos contrastes para descrever o perfil do vocabulário escrito dos nossos alunos. De acordo com Stubbs (2010, p. 23) apud Novodvorski (2016, p. 174), “palavras-chave são tipos de *icebergs*: ponteiros para objetos lexicais complexos, que representam as crenças e valores compartilhados de uma cultura”. Além disso, afirma que:

Esse autor destaca os significados sociais que expressam as palavras-chave, por fazerem parte do vocabulário de uma cultura e de uma sociedade, além do papel especial que desempenham nos textos, enquanto unidades de significado. Complementando, Scott (2010) observa que as palavras-chave funcionam como “ponteiros” para o pesquisador e que a



chavidade, nesse sentido, indica áreas que valeria mais a pena investigar, uma vez que essas palavras se tornam proeminentes por alguma razão que deveria ser analisada. (NOVODVORSKI, 2016, p. 174).

Nesse sentido, a lista de palavras-chave resultante do nosso contraste apontou para algumas palavras, demonstrando que, de fato, as *keywords* trazem muito da cultura, das crenças e dos valores compartilhados por aquele grupo, no caso, nossos alunos do Ensino Fundamental. Grande parte das palavras que aparecem no topo dessa lista está relacionada, principalmente, à temática dos textos e ao estilo dos autores. Isso pode ser observado no Quadro 28, a seguir:

Quadro 28 – Lista de palavras-chave do *corpus* de estudo em relação aos *corpora* de referência

Corpus de estudo/ Banco do Português	Corpus de estudo/ Lácio Web	Corpus de estudo/ Corpus Brasileiro
EU	EU	QUE
QUERO	QUERO	<b>EU</b>
<b>FILME</b>	<b>MINHA</b>	MAIS
<b>MINHA</b>	<b>FILME</b>	E
<b>PESSOAS</b>	<b>PESSOAS</b>	UM
<b>ELES</b>	<b>BRASIL</b>	UMA
FAMILIA	MEU	<b>PESSOAS</b>
<b>BRASIL</b>	<b>ELES</b>	EM
MAIS	FAMILIA	<b>BRASIL</b>
GOONIES	PRA	<b>ELES</b>
NÃO	CASA	<b>MINHA</b>
MEU	AMIGOS	NA
OQUE	MAIS	<b>FILME</b>
AMIGOS	QUERIA	TEM
COISAS	BOM	ELE
FUTURO	MEUS	MAS
TAMBEM	COISAS	SEM
QUERIA	TESOURO	DA
BOM	FUTURO	NÃO
PRA	TAMBEM	TODOS

Legenda: Negrito – palavras comuns entre os dois grupos

Fonte: Elaboração própria.

Como podemos observar no Quadro 28, com as 20 primeiras palavras da lista de *keywords*, algumas delas se repetem entre uma comparação e outra. São elas todas as palavras grifadas em negrito. Entre essas coincidências, estão o item “EU”, marca da primeira pessoa; o uso do pronome “MINHA”, também de primeira pessoa;

além das temáticas dos textos com maior concentração no *corpus*: “FILME” e “BRASIL”.

De acordo com Viana (2010, p. 65), “[...] as palavras-chave geralmente são de três tipos: temáticas, gramaticais e/ou identificativas”. O autor caracteriza esses três tipos, explicando que:

A primeira categoria abarca todas as palavras que apontam para o assunto abordado no *corpus* de estudo, as quais seriam inicialmente identificadas por um analista humano. [...] As palavras gramaticais, por serem geralmente empregadas em quantidades semelhantes em ambos os *corpora*, não aparecem no topo da lista de palavras-chave como ocorre numa lista de palavras regular. Contudo, o surgimento delas – com altos valores de chavidade – indica uma característica estilística do *corpus* estudado. [...] Finalmente, as palavras identificativas correspondem aos nomes próprios: por serem formas únicas de expressão, é esperado que apareçam entre as palavras-chave. (VIANA, 2010, p. 65).

Em nosso estudo, então, as palavras “FILME”, “BRASIL” e “PESSOAS” poderiam ser classificadas como palavras-chave temáticas; e “EU”, “ELE” e “MEU” como palavras-chave gramaticais. As primeiras fazem parte dos assuntos principais abordados nos textos dos alunos; as seguintes indicam o estilo da escrita desses alunos-redatores.

Quanto às palavras que Viana (2010) chama de identificativas, observamos que apenas no primeiro contraste, com o Banco de Português, apareceu um nome próprio entre as 20 palavras-chave de maior índice de chavidade. A palavra “GOONIES” faz parte do título de um dos filmes que fizeram parte das propostas de produção sobre filmes: *Os Goonies*, filme antigo, produzido por Steven Spielberg, em 1985. Nos outros contrastes, não encontramos palavras-chave identificativas entre as 20 primeiras.

Vale comentar também que os dois primeiros *corpora* do Quadro 28, Banco de Português e Lácio Web, apresentaram ainda mais proximidade nos resultados. Além das palavras em negrito, foram coincidentes também as formas: “QUERO”, “MEU”, “FAMÍLIA”, “AMIGOS”, “COISAS”, “QUERIA”, “FUTURO”, “BOM”. Ou seja, 15 das 20 palavras-chave se repetem entre os dois.

Outra questão que precisa ser salientada é a presença das palavras com grafias incorretas. Como já mencionado em alguns momentos deste estudo, optamos por não corrigir as palavras escritas pelos alunos, de maneira que o *corpus*

foi digitado conforme escrito originalmente, uma vez que, assim, representa o vocabulário e a especificidade desse público.

Dessa forma, considerando o universo de palavras-chave comum aos três *corpora* de referência (EU, MINHA, ELES, FILME, BRASIL, PESSOAS), vemos que não há repercussão do erro ortográfico sobre a identificação de especificidade. Na lista das primeiras 20 palavras-chave, percebemos apenas “FAMILIA” e “TAMBEM” sem acento e “OQUE” escrito junto. Então, podemos afirmar que a escolha por manter a escrita original dos alunos-autores não influenciou nos nossos resultados, já que foram poucas as ocorrências na amostra estudada.

Outra ocorrência que nos chamou atenção foi o uso do “PRA”, que aparece nos dois primeiros contrastes, marcando a presença da oralidade na escrita. Acreditamos que tal elemento também caracteriza o estilo de escrita desses redatores. O seu uso no *corpus* corresponde a 70 ocorrências, valendo observar que, em contraste, “QUE”, por exemplo, exibe 1.211 ocorrências.

Estudos sobre palavras-chave vêm sendo desenvolvidos no âmbito da Linguística de Corpus, pois proporcionam “[...] uma seleção dos itens lexicais de seu *corpus* de estudo que são estatisticamente mais distintivos” (BERBER SARDINHA, 2006, p. 02). Essas pesquisas buscam autoria, estilo e marcas, opiniões e ideologias, formação de glossários, entre outros. A seguir, apresentamos a síntese dos resultados obtidos ao longo de nossa pesquisa e apresentados neste capítulo.

## 5.6 SÍNTESE DOS RESULTADOS

A partir de nossos quatro estudos piloto, revisitamos nossas análises, nossos materiais e ajustamos nossa metodologia. Isso possibilitou afinar a pesquisa como um todo. Chegamos, então, ao experimento principal, com o todo de nosso *corpus* de estudo, com métodos de investigação mais refinados.

Os resultados também foram se modificando à medida que ampliamos a quantidade de redações escolares sob exame. Dessa forma, na Figura 7, a seguir, é possível visualizar o caminho percorrido pelos estudos parciais, bem como pelo experimento principal. Fica ilustrada a quantidade de textos analisados, assim como o número de *tokens*, de *types* e a riqueza lexical nos diferentes momentos da pesquisa:

Figura 7 – Informações gerais – estudos piloto e experimento principal

Escola Municipal	Estudo Piloto I	Estudo Piloto II	Estudo Piloto III	Experimento Principal
<i>Tokens</i>	836	3.863	8.382	14.735
<i>Types</i>	350	1.076	1.731	2.444
Riqueza Lexical	41,9%	27,8%	20,6%	16,5%

Quantidade de textos →	↑ 4 ↓	↑ 18 ↓	↑ 47 ↓	↑ 77 ↓
------------------------	-------------	--------------	--------------	--------------

Escola Estadual	Estudo Piloto I	Estudo Piloto II	Estudo Piloto III	Experimento Principal
<i>Tokens</i>	729	3.902	9.925	15.732
<i>Types</i>	306	1.167	2.196	2.609
Riqueza Lexical	41,8%	29,9%	22,2%	16,5%

Fonte: Elaboração própria.

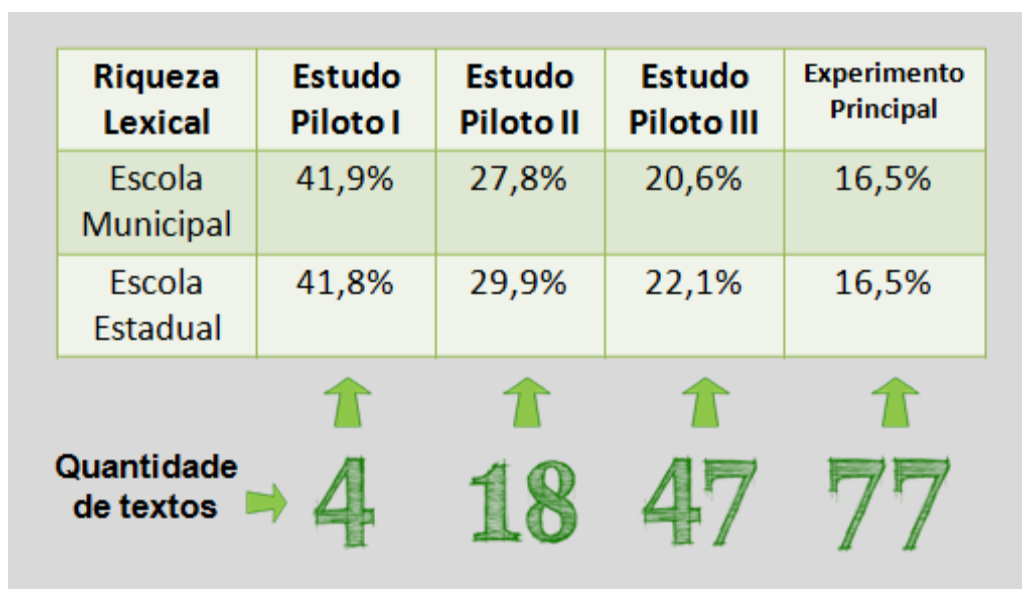
Conforme a Figura 7, à medida que aumentamos o número de redações analisadas, o número de *tokens* e de *types*, por óbvio, também aumentou. É importante salientar, porém, que esses valores sempre foram maiores na escola estadual. Somente no primeiro estudo, com apenas quatro redações de cada grupo, os estudantes da escola municipal escreveram mais e com mais palavras diferentes.

Considerando o quadro geral da pesquisa, concluímos que os alunos da escola estadual tendem a escrever mais. Além disso, empregam maior diversidade de palavras nas suas redações.

Com relação à riqueza lexical, essa medida foi diminuindo à medida que a quantidade de textos (e, por consequência, de *tokens*) foi aumentando, algo que também já era esperado. Afinal, essa é uma tendência geral da medida de riqueza lexical. Ainda assim, a riqueza lexical manteve-se sempre um pouco mais alta nos textos da escola estadual. Mesmo naquele único estudo em que o grupo da escola municipal apresentou número maior de *tokens* e de *types*, a diferença de riqueza

ficou em torno de 0,01%, ou seja, não houve diferença relevante. Isso é o que está sintetizado na Figura 8 a seguir:

Figura 8 – Riqueza Lexical – estudos piloto e experimento principal



Fonte: Elaboração própria.

Como é possível perceber na Figura 8, no experimento principal é que essa riqueza surpreendeu ao apresentar exatamente o mesmo valor de 16,5% de variedade vocabular, independente da escola e das quantidades de *tokens* e de *types*. Nesse sentido, deduzimos que a variação vocabular das duas escolas analisadas tende a ser muito semelhante, para não dizer igual, no caso do experimento principal. Apesar de quantidades diferentes de palavras, na proporção, os estudantes variaram o seu vocabulário na mesma medida.

Concluimos, também, com relação ao grupo das palavras mais frequentes, que a proporção entre o repertório lexical e gramatical tende a ficar bem próxima nos textos das duas escolas. Os alunos utilizam em seu vocabulário escrito cerca de 60% de palavras gramaticais, frente a um universo de cerca de 40% de palavras lexicais. Isso foi observado nos estudos piloto III e IV e no experimento principal.

Ressaltamos que os resultados do estudo piloto IV não aparecem no Quadro 28, nem na Figura 8, anteriores, porque esse estudo não contrastou as redações dos dois grupos de escolas, assim como todos os que ali aparecem. Foram comparados todos os textos apenas por ano/ciclo de aprendizagem. No contraste do estudo piloto IV, não há diferença considerável nas quantidades de *tokens* ou de

*types*, ao longo do período escolar dos alunos, que vai do 6º ao 8º ano. Esse resultado levou a uma riqueza lexical bastante próxima também: 20,8% no grupo do 6º ano/B30 e 17,4% no grupo do 8º ano/C20. Nesse sentido, tais dados ajudaram a comprovar semelhança também entre os textos dos alunos dos diferentes anos/ciclos pesquisados.

Já com relação ao conjunto de palavras que compõem a camada *hapax*, as variações foram maiores a cada estudo piloto e, sobretudo, no experimento principal. Na Figura 9, a seguir, podemos visualizar a porcentagem de coincidência semântica entre os *hapax* das letras A e P ao longo dos estudos:

Figura 9 – Porcentagem de *hapax* coincidente – estudos piloto e experimento principal

	<b>Estudo Piloto I</b>	<b>Estudo Piloto II</b>	<b>Estudo Piloto III</b>	<b>Experimento Principal</b>
<b><i>Hapax coincidente</i></b>	38,6%	58,3%	45,5% / 7,5%	11,4%

Fonte: Elaboração própria.

O conjunto de palavras da camada *hapax* revela coincidência expressiva nos três primeiros estudos. Isso é observado principalmente no segundo estudo piloto, quando nossas listas ainda não estavam lematizadas. A partir do estudo III, quando lematizamos as listas, percebemos que: a) a quantidade de palavras nessa camada diminuiu consideravelmente; b) a coincidência no vocabulário escrito utilizado pelos alunos dos dois grupos também ficou menor.

No terceiro estudo, por exemplo, em que a coincidência de palavras do *hapax* entre os textos dos dois grupos era de 45,5%, ao lematizarmos, passou para apenas 7,5%. Essa porcentagem baixa se manteve no estudo piloto IV, que não está no quadro, por relacionar os dois diferentes anos/ciclos analisados e não as escolas. No estudo IV, apenas 12% de palavras da camada *hapax* dessas das letras A e P coincidiram, independente da idade e do ano/ciclo dos alunos.

Em nosso experimento principal, a semelhança entre os itens do *hapax* dos dois grupos também foi baixa: 11,4%. Em função disso, apesar de todos os outros itens terem aproximado o perfil de vocabulário escrito dos dois grupos, concluímos que está no *hapax* a maior diferença entre esses perfis. Essa camada foi justamente

a que defendemos desde o início que seria a mais promissora. Esperávamos que nessa camada estivessem os elementos de maior particularização dos perfis lexicais.

Ainda que os alunos aparentemente tenham vocabulário escrito (com relação tanto à gramática, quanto ao léxico) mais frequente bastante próximo, quando analisamos o vocabulário específico de cada grupo, via *hapax*, percebemos grande distanciamento nas preferências de ambos. Tanto em relação às palavras iniciadas pela vogal A, quanto às iniciadas pela consoante P, as especificidades de vocabulário de cada grupo pouco coincidiram. Nos dois grupos de palavras, coincidiram em apenas 9% e 12% dos itens. Esses dados mostram que, de fato, o *hapax* é o que mais diferencia os perfis vocabulares estudados.

A fim de verificar em que medida nosso *corpus* de estudo, como um todo, se aproxima (ou não) de um *corpus* de língua geral, também o contrastamos com o CorPop. Conforme já mencionado, o CorPop foi utilizado como *corpus* de referência do português popular brasileiro escrito culto, ou seja, “[...] um texto compatível com o nível de letramento médio da maioria dos brasileiros adultos [...]” (PASQUALINI, 2018, p. 14). Dessa comparação concluímos que quase 70% das palavras do nosso *corpus* de estudo está presente no CorPop. Então, é possível afirmar que o perfil de vocabulário escrito dos alunos tende a se aproximar de um perfil de português popular escrito culto.

Considerando a especificidade do CorPop, optamos por também analisar as palavras-chave do nosso *corpus* de estudo como um todo, frente a outros três *corpora* de referência da Língua Portuguesa: Banco de Português, Lácio Web e Corpus Brasileiro. Esse contraste complementou os resultados que já havíamos obtido, ajudando na descrição do perfil vocabular desses alunos. As *keywords* nos mostraram as temáticas principais do *corpus* e as preferências dos alunos-redatores.

A seguir, no próximo capítulo, retomamos nossas questões de pesquisa e nossas hipóteses, a fim de discuti-las de acordo com nossos experimentos.

## 6 RETOMADA DAS QUESTÕES DE PESQUISA E HIPÓTESES

Conforme detalhado no capítulo anterior, nosso experimento descritivo com o vocabulário das redações dos alunos do Ensino Fundamental das duas escolas públicas nos permitiu responder nossas questões de pesquisa. Além disso, possibilitou que pudéssemos ponderar sobre as nossas hipóteses. Essas ponderações são apresentadas a seguir, junto das discussões realizadas em torno de cada uma, a partir dos resultados obtidos.

### 6.1 QUESTÕES DE PESQUISA

Levantamos três questões principais. Elas foram sendo parcialmente respondidas conforme os resultados dos nossos estudos piloto. Com o nosso experimento principal, buscamos construir respostas mais consistentes. Afinal, nessa última fase, lidamos com todos os textos de uma só vez. Vamos às questões:

#### **a) Os alunos dessas duas redes de ensino público apresentam diferentes perfis de vocabulário escrito?**

**Resposta:** Não. O perfil é semelhante, tendo sido percebido por meio da análise de características como: número de *tokens* e de *types*, frequência de palavras (gramaticais e lexicais) e riqueza lexical. Dessa forma, do ponto de vista léxico-estatístico, não podemos dizer que há diferentes perfis de vocabulário escrito entre os dois principais grupos estudados.

Com relação ao número de *tokens*, percebemos que os alunos da escola estadual escreveram aproximadamente 1.000 palavras a mais nos seus textos. Esse número, descontextualizado, parece significativo, porém, no universo do total dos 77 textos analisados de cada grupo, são cerca de 13 palavras a mais por texto, apenas. Em relação aos *types*, o resultado é ainda mais próximo. Os textos da escola estadual apresentaram 165 *types* a mais do que os da escola municipal, entretanto, quando calculamos a média, foram somente 6 *types* a mais por texto.

O conjunto de redações das duas escolas também apresentou o uso de palavras gramaticais semelhantes, quando consideradas as mais frequentes. Cerca de 87% das palavras gramaticais presentes entre as 50 palavras mais frequentes coincidiram, indicando que os estudantes tendem a utilizar, com preferência, os mesmos recursos gramaticais. Podemos citar o uso do “mas” e do “depois”,



utilizados com preferência em relação ao “porém” e ao “após”, por exemplo. Esse indicativo revela uma coincidência de repertório gramatical comum aos dois grupos, como é possível perceber nos trechos abaixo, retirados das redações escolares pesquisadas:

Quadro 29 – Exemplos de contextos das palavras gramaticais mais usadas

Escola Municipal – GRUPO A	Escola Estadual – GRUPO B
A família do meu pai não gostou muito da ideia de fazer a cirurgia <b>mas</b> minha mãe já estava decidida	Temos altos e baixos <b>mas</b> nos mantemos sempre unidos independente do que aconteça.
<b>Depois</b> de algumas horas de aula voltamos a nossa casa.	<b>Depois</b> de um tempo virei o melhor amigo do Cristiano Ronaldo [...]

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

A partir do Quadro 29, percebemos que, além do repertório de palavras gramaticais ser muito semelhante, suas construções frasais/oracionais também são próximas. Na segunda frase de cada coluna, por exemplo, ambas as construções combinaram o “depois” com o “de” e deixaram bem marcada a ideia de tempo com os usos de “algumas horas” e “um tempo”. Dessa forma, a aproximação entre os grupos fica bastante visível.

Além da proximidade entre os grupos, esses resultados vão ao encontro do que afirma Biderman (1998), que, ao analisar seu *Dicionário de frequências do português do Brasil*, concluiu que, entre as palavras de altíssima frequência estão, sobretudo, as palavras instrumentais, em nosso estudo, chamadas de palavras gramaticais.

Além da coincidência de 87% no repertório de palavras gramaticais, vale ponderar que se observa um ponto de divergência, menos acentuado, corresponde ao emprego dos itens “ou”, “porque”, “pra”, “tudo”, “me”, “das”, “elas” (GRUPO A – 7 itens) e “eles”, “como”, “sua”, “seu” (GRUPO B – 4 itens).

As palavras lexicais mais utilizadas coincidiram em menor quantidade entre os textos dos dois grupos. Entretanto, chamou-nos atenção que itens lexicais que não migraram dos textos de apoio (como, por exemplo, “pessoa”, “casa”, “amigo” e “professor”) apareceram nos resultados dos dois grupos de redações analisados. Sabemos que podem ser itens usuais do universo dos alunos, entretanto, chamou atenção serem utilizados com maior frequência pelos dois grupos. Isso é o que ilustram os trechos do Quadro 30:

Quadro 30 – Exemplos de contextos das palavras lexicais mais usadas

Escola Municipal – GRUPO A	Escola Estadual – GRUPO B
Se eu criasse um filme as primeiras <b>pessoas</b> que eu iria por no filme era minha mãe, meu pai e meus irmãos tendo tudo que ele sonharem.	Então eu quero um Brasil com mais educação, os pagamentos certos das <b>pessoas</b> que necessitam do dinheiro e é isso que eu quero para o Brasil.
[...] eu agradeço a deus por acada dia ter a família e <b>amigos</b> ao meu lado.	Meus atores seriam os que mais se parecem com meus <b>amigos</b> /conhecidos atuais; não pela beleza, isso é só um bônus.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

Como podemos ver em alguns trechos das redações presentes no Quadro 30, essa proximidade acontece mesmo quando as temáticas das redações são diferentes. A família e os amigos são temáticas recorrentes em todas as propostas, mostrando a preocupação dos redatores com o bem-estar e a realização dessas pessoas que consideram importantes nas suas vidas. Dessa forma, os assuntos prioritários desses estudantes tendem a ser os mesmos, independente das propostas de produção serem variadas e do nosso *corpus* de estudo ser heterogêneo nesse quesito.

Conforme discutido no Capítulo 2 e de acordo com Biderman (1998b), as pessoas tendem a preferir e, portanto, usar com mais frequência, aquele vocabulário que vai ao encontro das suas experiências e vivências. Em função disso, encontramos essa semelhança no repertório de palavras lexicais de ambos os grupos de textos. Ainda que as propostas temáticas tenham sido diferentes, as experiências e vivências desses estudantes, pertencentes a comunidades tão próximas, são semelhantes.

Além disso, a variedade vocabular encontrada nas redações dos dois grupos foi exatamente a mesma (quando consideradas as mesmas quantidades de texto, o todo do *corpus* de estudo, independente das propostas variadas): em torno de 16,5%. Tal fato aproxima e, desta vez, até mesmo iguala os dois grupos nesse quesito. Isso também demonstra proximidade no perfil de vocabulário escrito desses alunos. Os dois trechos abaixo exemplificam uma outra faceta desta mesma situação:

Tabela 22 – Exemplos de riqueza lexical

Escola Municipal – GRUPO A	Escola Estadual – GRUPO B
Bom, o Brasil que eu quero para o futuro é um Brasil sem imposto, sem crimes e com mais educação.	Sei que eu não estou gravando um video de verdade mais é o que eu acho que deveria mudar.
Riqueza Lexical: 85%	Riqueza Lexical: 84,2%

Fonte: Elaboração própria.

Na Tabela 22, os exemplos são apenas uma frase de cada um dos dois principais grupos pesquisados. Ainda assim, mesmo em um pequeno trecho, percebemos que a variedade vocabular é muito próxima, ficando visível também essa semelhança entre os perfis vocabulares.

Em função de todos os elementos descritos, podemos afirmar que os alunos das duas redes públicas de ensino não apresentam diferenças significativas de perfis de vocabulário escrito. A semelhança de vocabulário entre esses dois grupos de redatores fica evidente quando analisadas suas produções escritas.

#### **b) Quais elementos de natureza estatística permitem caracterizar o vocabulário escrito dos dois públicos?**

**Resposta:** Os elementos de natureza estatística que nos permitiram, nesta pesquisa, caracterizar o vocabulário escrito dos dois públicos, foram: número de *tokens* e de *types*; riqueza lexical; proporção e análise de palavras gramaticais e lexicais mais frequentes; itens de ocorrência única, da camada *hapax legomena*; *keywords*.

Estudos como Biderman (1998) e Berber Sardinha (2004), por exemplo, comprovaram que a frequência das palavras é importante, pois revela o padrão de uso da língua, auxiliando na sua caracterização, bem como no entendimento da língua como um todo. Além disso, as frequências são muito usadas na confecção de materiais didáticos para o ensino de línguas. Então, como nosso objetivo também era o de produzir material didático que pudesse auxiliar o professor em sala de aula, a análise das palavras mais frequentes no *corpus* de estudo foi de grande valia, pois auxiliou não só na caracterização dos perfis de vocabulário escrito, como também na confecção das atividades.

As semelhanças no perfil de vocabulário escrito dos dois principais grupos foram verificadas, conforme já mencionado na primeira questão, por meio da análise

das quantidades de *tokens* e de *types*, das palavras mais frequentes presentes nas redações desses alunos e da proporção entre os itens gramaticais e os itens lexicais. Além desses, a riqueza lexical também foi traço léxico-estatístico utilizado para caracterizar o vocabulário escrito desses alunos.

A partir da análise desses itens anteriormente citados, ficou comprovado que há semelhança entre os dois repertórios analisados, conforme a amostra de textos que fixamos para estudo. Os resultados mostraram aproximação entre os usos e as preferências dos estudantes escritores, independente das propostas, conforme já exemplificado na questão anterior.

Essa proximidade visualizada entre os repertórios vocabulares dos dois grupos de textos analisados corrobora com os achados de Nascimento e Isquierdo (2003), que afirmam que talvez possam existir palavras “multiuso”, que aparecem em qualquer tipo de texto. Nosso estudo, com redações escolares, mesmo que de temáticas variadas, contribui com os resultados já encontrados pelas colegas pesquisadoras.

Dos itens de natureza estatística descritiva que nos ajudaram a caracterizar o *corpus* de estudo, o único que diferenciou os dois grupos pesquisados foi a análise dos *hapax*, itens de frequência única. Esses itens são detalhados e especificados na próxima questão, que trata justamente da diferença entre os dois grupos de textos.

Em função dessa distinção, investigamos também as *keywords* de nosso *corpus* de estudo. De acordo com o explicado no Capítulo 2, sessão 2.2.10, essas unidades correspondem a especificidades do *corpus* de redações em relação aos *corpora* de referência, que representaram, nessa pesquisa, o todo da Língua Portuguesa. A partir do índice de chaticidade das palavras encontradas como *keywords*, foi possível caracterizar o *corpus* de estudo como um todo. Percebemos a preferência dos alunos pelo uso da primeira pessoa, com os pronomes “EU” e “MEU”, além das marcas da oralidade na escrita, com o uso do “PRA”. Isso é o que vemos no Quadro 31, a seguir:

Quadro 31 – Exemplos de contextos de palavras-chave

<p>Talvez um dia <b>eu</b> vá pra Disney e isso com certeza <b>eu</b> não vou deixar passar, mais se <b>eu</b> não for contarei sobre minhas outras viagens, contarei também o dia em que eu sai de Belém e vim morar no Rio Grande do Sul (Porto Alegre).</p> <p><b>Eu</b> estava indo falar com com <b>meus</b> pais quando minha bicicleta escorregou e cai com tudo no chão.</p>
<p><b>Meu</b> Brasil para o futuro (em <b>meu</b> pensamento) é um Brasil com mais segurança, sem ladrões, mais policiais, com mais saúde, mais educação, [...]</p> <p><b>Meu</b> pai seria o ator principal, <b>eu</b> eliminaria muitas coisas tristes, pois passei muitas.</p>
<p>Gostei e recomendo <b>pra</b> todos.</p> <p><b>Meus</b> pais sempre foram bons <b>pra</b> mim, <b>eu</b> admiro muito eles.</p>

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

De acordo com o Quadro 31, acima, percebemos que, em muitos dos trechos aparecem, inclusive na mesma frase, as três formas: “eu”, “meu” e “pra”. Isso mostra como essas palavras realmente são chave e têm recorrência no todo do *corpus* de estudo. Tais especificidades podem estar atreladas também a um tipo de gênero textual e discursivo específico, no caso, a redação escolar do Ensino Fundamental.

É importante ressaltar que, de acordo com Berber Sardinha (2009), palavras-chave não são o mesmo que palavras importantes. Enquanto um analista humano poderia selecioná-las de acordo com sua relevância para o texto, a ferramenta utiliza critérios estatísticos e quantitativos de diferença entre o *corpus* de estudo e o *corpus* de contraste. Ou seja, essas palavras-chave não são sempre as mesmas, elas são relativas ao *corpus* de referência utilizado. Em nosso caso, as palavras-chave que discutimos em nossos exemplos apareceram como chave na maioria dos três contrastes realizados.

As temáticas dos textos em maior quantidade no *corpus* também apareceram entre os mais altos índices de chavidade: “FILME” e “BRASIL”. Do mesmo modo, “EU” e “MEU” também podem ser considerados como temas dos textos. Assim, acreditamos que o índice de chavidade, bem como a lista de palavras-chave também contribuíram para caracterizar o universo vocabular e temático dos textos sob exame e para distinguir o *corpus* de estudo frente ao todo da língua portuguesa.

### c) Quais elementos de natureza estatística permitem distinguir o vocabulário escrito de um grupo em relação ao outro?

**Resposta:** As diferenças entre os perfis de vocabulário escrito desses alunos foram verificadas, principalmente, a partir da análise de itens de frequência única, chamados de *hapax legomena*. Foi a partir da análise desses itens que percebemos

que havia distinções no repertório vocabular desses alunos. Se as palavras do grupo *hapax* são capazes de nos mostrar identidade ou autoria de um texto, já que correspondem a um repertório de palavras de escolha de uso mais livre, no nosso caso, revelaram um traço que foi capaz de relativizar a homogeneidade percebida nas questões anteriores.

De acordo com o exposto no Capítulo 2, Berber Sardinha (2000) assegura que as formas de frequência um são a maioria em qualquer *corpus*. Então, para analisarmos essa camada, recortamos uma amostra do total dos resultados de 2.308 itens do grupo *hapax*. A amostra compreende os *hapax* iniciados pela vogal A e pela consoante P. Lembramos que, conforme já explicado no Capítulo 5, a escolha pelos itens iniciados por essas letras foi realizada baseada em metodologia usada em estudos lexicográficos. Nesses estudos, avalia-se a cobertura de repertório vocabular em dicionários por amostragem. Isso é o que vemos no trabalho de Damim (2005) e em materiais do PNLD sobre avaliação de Dicionários (BRASIL, 2012). A seleção dessas letras justifica-se em função da abrangência e da diversidade de itens correspondentes.

Percebemos, então, que as quantidades de *hapax* eram semelhantes nos dois grupos: 54-57 / 42-47, respectivamente. Entretanto, ao analisarmos quais itens estavam presentes em cada grupo, a semelhança foi mínima.

Na letra A, apenas os verbos “aceitar”, “agradecer”, “apresentar”, “assustar” e “atirar” coincidiram. Na letra P, coincidiram itens de outras classes gramaticais, semelhantes quando levadas em conta suas características semânticas: “padaria”, “perigo”, “piscina”, “play”, “povo”, “presidente” e “provavelmente”. Ou seja, houve apenas 9% de itens coincidentes no *hapax* dos dois grupos de textos na letra A, e 12% na letra P. Isso significa que, quando o redator pode escolher livremente as palavras, cada grupo de estudantes faz escolhas diferentes. Assim, a análise dos itens específicos de cada grupo demonstrou que, de fato, existem diferenças entre o vocabulário escrito dos alunos das duas redes. Essas diferenças são detalhes em meio a um conjunto de coincidências.

**d) Como são os perfis de vocabulário escrito dos diferentes grupos de redatores envolvidos na amostra?**

**Resposta:** Os redatores envolvidos na amostra pesquisada foram categorizados em quatro diferentes grupos, de acordo com sua escola/rede ou conforme seu ano/ciclo. O Quadro 32, a seguir, mostra essa composição:

Quadro 32 – Composição dos grupos de alunos pesquisados

<b>Grupo</b>	<b>Composição</b>
Grupo A	Alunos da Escola Municipal
Grupo B	Alunos da Escola Estadual
Grupo C	Alunos do 6º ano/B30 (das duas redes)
Grupo D	Alunos do 8º ano/C20 (das duas redes)

Fonte: Elaboração própria.

GRUPO A: os textos apresentam menos quantidade de palavras, bem como menos palavras diferentes também, quando comparados aos resultados do Grupo B. Quanto à riqueza lexical, essa medida se igualou entre os textos dos Grupos A e B, apesar dos resultados diferentes de *tokens* e de *types*.

Com relação às palavras gramaticais, os resultados do Grupo A apresentam muitas semelhanças com todos os outros grupos, exibindo grande quantidade dessas palavras entre as mais frequentes. Destacamos, como singularidade, a preferência de uso do conetivo “porque”, que não apareceu entre os recorrentes do Grupo B. Isso é o que vemos nos exemplos do Quadro 33, a seguir:

Quadro 33 – Exemplos de uso do “porque” – Grupo A

No meu filme teria coisas ruins <b>porque</b> se é o filme da minha vida não pode ter so coisas boas [...]
[...] nós temos que dar valor para as coisas que nós temos, <b>porque</b> se não darmos valor nada vai dar certo [...]
Pra mim, voltar de ferias é meio chato <b>porque</b> você não pode mais ficar em casa pra ver serie e dormir tarde.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

O uso preferencial do “porque” pode estar demonstrando uma preocupação dos alunos em se explicar, em esclarecer suas opiniões e os motivos pelos quais defendem suas ideias. Além disso, eles parecem utilizar com frequência o “porque” como forma de justificar seu posicionamento, conforme orientação das propostas de produção (ANEXO II): “não copie as ideias do(a) autor(a), seja criativo e justifique suas opiniões”.

Já em relação às palavras lexicais, algumas estão diretamente relacionadas às propostas e temáticas das produções. Entretanto, outras particularizam o universo dos escritores deste grupo. São elas: “dia”, “novo”, “passar”, “ver”, “ajudar” e “saber”. Os trechos a seguir ilustram alguns desses usos nas redações dos alunos do Grupo A:

Quadro 34 – Exemplos de uso de palavras lexicais – Grupo A

Até que não é tão ruim assim o primeiro <b>dia</b> .
No <b>dia</b> eles estavam todos se aprontando para comemorar o ano <b>novo</b> , aí minha mãe começou a passar mal [...]
Uma das lições foi que temos que <b>ajudar</b> ao próximo [...]

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

Mais uma vez, percebemos que os assuntos de preferência dos alunos das duas diferentes realidades são os mesmos, independente das temáticas serem variadas. Até mesmo dentro do grupo exclusivo de determinada rede/escola, percebemos que o vocabulário escrito empregado é bastante parecido, embora tenhamos redatores de diferentes anos e faixas etárias escrevendo sobre temáticas diversas.

GRUPO B: os textos desse grupo apresentam, em média, número maior de palavras (*tokens*) do que os textos do Grupo A. Isso obviamente revela que os alunos deste Grupo B produzem textos mais longos. Outra especificidade é em relação ao repertório de palavras diferentes (*types*), levemente maior no grupo B. Apesar desses resultados distintivos, a riqueza lexical é um fator que não distingue um grupo do outro, como já mencionado.

Conforme o esperado, já que é uma regra geral da distribuição estatística de vocabulário nas línguas românicas, as palavras de maior frequência nesse grupo também foram as gramaticais. A especificidade fica por conta da preferência de uso do conetivo “como”, conforme visualizamos no Quadro 35, a seguir, não muito utilizado pelo Grupo A:

Quadro 35 – Exemplos de uso do “como” – Grupo B

Com certeza a sua realidade não era <b>como</b> o mundo era realmente.
Mas <b>como</b> diz a minha mãe: nem tudo é perfeito [...]
Sou uma menina normal <b>como</b> todas as outras, mas não me sinto em padrão de beleza muito bom [...]

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.



A partir dos trechos de exemplo mencionados acima, percebemos que o uso do “como” preferencial dos alunos é no seu sentido comparativo. Eles frequentemente comparam realidades, modos de dizer, pessoas. Então, podemos afirmar que os redatores desse grupo tendem a comparar com frequência o que expõem.

Já em relação às palavras lexicais, algumas oriundas dos textos de apoio e das propostas de produção também apareceram com regularidade. Ainda assim, o que os diferenciou foram suas preferências de uso por itens lexicais como: “comer”, “pai”, “coisa”, “achar”, “estar”, “rua” e “conseguir”, itens menos empregados pelos estudantes do Grupo A. No Quadro 36, a seguir, estão alguns trechos que ilustram o uso dessas palavras nos textos dos alunos do Grupo B:

Quadro 36 – Exemplos de palavras lexicais – Grupo B

As pessoas que eu destacaria seria eu, minha mãe, meu <b>pai</b> , e minha irmã, e o meu cachorro e meus avós.
O Brasil de hoje está uma vergonha total, tem muita violência, não posso sair na <b>rua</b> totalmente tranquila [...]
[...] os meninos ficam pensando no que fazer para <b>conseguir</b> o dinheiro.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

De acordo com o Quadro 36, percebemos a preocupação dos redatores do Grupo B com a família, que teria destaque no filme da vida deles. Essa temática é abordada em todas as propostas e aparece em muitos dos quadros com trechos de exemplo, demonstrando a importância que a família tem para os redatores. Além disso, a violência é um assunto que também aparece com frequência nos textos dos alunos, afinal, está cada dia mais presente na vida das pessoas, preocupando a todos, inclusive jovens e adolescentes.

Por fim, o verbo “conseguir”, apontando outra questão em destaque na vida deles: “conseguir” dinheiro. Muitos dos alunos já estão em idade de iniciar em programas de estágio ou jovem aprendiz, buscando ter o próprio dinheiro. Então, é natural o interesse deles por esse assunto também.

Os Grupos C e D somente foram pesquisados no estudo piloto IV, quando contrastamos os dois diferentes anos envolvidos na pesquisa. Julgamos importante lembrar que as quantidades de textos por ano contidas em nosso acervo eram

diferentes. Dessa forma, fizemos uma proporção para podermos equiparar os resultados, comparando-os de igual para igual, no mesmo patamar.

GRUPO C: os alunos escrevem menos em suas redações e com menos quantidade de palavras diferentes que os estudantes do Grupo D. Entretanto, quando analisamos a riqueza lexical, em proporção, percebemos que a variedade desse vocabulário é bastante semelhante à do grupo D, em torno de 3% mais variado, somente.

Considerando as 50 palavras mais frequentes empregadas nos textos do Grupo C, ficou constatado que 60% delas são palavras gramaticais. Essa tendência pode ser considerada uma especificidade, pois a mesma observação em um texto de outro gênero pode ser diferente. No romance *A casa das sete mulheres*, de Letícia Vierzchowski, escrito em 2002, temos um universo de cerca de 150 mil palavras. Considerando o conjunto das 50 palavras mais frequentes, 74% são palavras gramaticais. Já num universo de 70 notícias de um jornal popular de Salvador-BA, o jornal *Massa!*, essa mesma proporção atingiu 70% de palavras gramaticais. Esses resultados indicam que o universo de formas gramaticais das redações escolares tende a ser menor, supomos que por terem sido escritas por alunos de Ensino Fundamental. As análises dos outros textos apontaram para repertórios gramaticais maiores, provavelmente porque foram escritos por adultos, profissionais e experientes. Tais fatos renderiam, inclusive, um novo estudo, mais aprofundado e específico, em que se poderia relacionar os dados obtidos com dados de outros tipos de textos de aprendizes, como de exames de proficiência (EVERS, 2013) ou redações de vestibular (EVERS, 2018).

Dentre os usos gramaticais específicos dos textos do Grupo C, destacamos as palavras “sem”, “isso”, “porque”, “melhor”, “menos”, “também”, “esse” e “hoje”, itens não relacionados entre os recorrentes do Grupo D. No Quadro 37, a seguir, estão alguns exemplos desses usos:

Quadro 37 – Exemplos de palavras gramaticais – Grupo C

[...] o brasil que eu quero é um Brasil <b>sem</b> poluição, <b>sem</b> preconceito, <b>sem</b> corrupção, com paz e amor.
<b>Esse</b> é o país do meu sonho e o que eu quero viver!
Olá Porto Alegre, <b>hoje</b> eu vou falar um pouquinho do Brasil que eu quero para o futuro, vou começar.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

A partir dos usos do Grupo C, conforme o Quadro 37 acima, percebemos que a construção dos alunos é bastante próxima do texto de apoio e, principalmente, dos vídeos da campanha da Rede Globo mostrados na televisão. Além disso, destacamos a repetição da palavra “sem”, por três vezes na mesma frase. Acreditamos que isso representa a marcação incisiva do que o estudante realmente não quer que tenha mais no seu país.

Entre as palavras lexicais, novamente encontramos elementos dos textos de apoio entre os mais frequentes. Ainda assim, houve singularidades que apareceram entre as palavras lexicais mais frequentes deste Grupo C, sem ligação com a temática das produções. São elas: “rua”, “ir”, “lixo”, “poder” e “parar”. Alguns exemplos de uso estão no Quadro 38, a seguir:

Quadro 38 – Exemplos de palavras lexicais – Grupo C

Que os jovem possam <b>ir</b> a escola com segurança [...]
Eu quero que as ruas de porto Alegre tenham menos <b>lixo</b> no chão.
Desejo <b>poder</b> sair á rua para brincar com a certeza de que voltarei para a casa em segurança.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

O Quadro 38 aponta assuntos importantes para os jovens quando questionados sobre o que querem para o seu país no futuro, à medida que crescem e se tornam adultos. Conforme já mencionado, um dos assuntos recorrentes nas redações escolares é a violência presente em nosso país. Mais uma vez esse assunto apareceu, demonstrando a preocupação desses adolescentes com a falta de segurança. Além disso, outro assunto recorrente nas redações dessa proposta é o lixo, a sujeira e a falta de educação das pessoas com relação a isso. Nesse sentido, acreditamos que os alunos de 6º ano/B30 têm opinião e senso crítico sobre o que acontece ao seu redor, expondo isso nas suas produções.

GRUPO D: os alunos escrevem mais em seus textos, com maior número de *tokens* e de *types*, quando comparados aos estudantes do Grupo C. Acreditamos que isso é algo natural, visto que são estudantes geralmente mais velhos e com dois anos a mais de estudo. Ainda assim, conforme já mencionado, em termos proporcionais, a riqueza lexical presente nesses textos é pouco menor que a do Grupo C.

As palavras gramaticais persistem como as mais frequentes em maior quantidade. Destacamos o uso de “quando”, “depois”, “tudo”, “outro”, “como”, “até”,

“pelo” e “então”, itens que não aparecem entre os recorrentes do Grupo C. Alguns trechos em que essas palavras aparecem estão no Quadro 39, a seguir:

Quadro 39 – Exemplos de palavras gramaticais – Grupo D

<b>Quando</b> terminou o serviço foi entregar a camiseta a dona da loja, <b>quando</b> chegou lá com a camiseta na mão a dona da loja disse a ela “Pegue Jennifer, esse é meu presente para você.
<b>Depois</b> de um tempo, minha mãe receberia uma proposta de emprego no Canadá e eu me mudaria para lá.
Os meninos que estavam na busca do tesouro tiveram que passar por diversas armadilhas <b>até</b> chegar em um ponto que os bandidos chegaram perto deles.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

De acordo com o Quadro 39, visualizamos que, apesar das temáticas do 8º ano/C20 serem diversas, a questão temporal aparece com frequência nos variados temas. Embora escolhidas ao acaso para exemplificar os usos gramaticais dos textos do Grupo D, em todos os trechos as palavras grifadas estão, de alguma forma, relacionadas ao tempo. Identificamos, com isso, que os redatores desse grupo têm uma preferência por deixar o tempo bem demarcado em suas redações.

Entre o repertório de palavras lexicais, encontramos as palavras advindas dos textos de apoio e das propostas de produção novamente. Outros elementos encontrados, não relacionados às propostas, foram: “amigo”, “pai” e “achar”. Esses itens particularizam o Grupo D quanto às suas preferências de escrita e de assuntos, conforme visualizamos no Quadro 40, a seguir:

Quadro 40 – Exemplos de palavras lexicais – Grupo D

Eu seria uma adolescente normal, sem muitos <b>amigos</b> , mas os que eu teria seriam os melhores.
Hoje meus pais são separados, meu irmão mora com meu <b>pai</b> e eu com minha mãe, nos vemos tão pouco.
No primeiro dia é muita confusão, você tem que chegar na escola, passar pela multidão depois <b>achar</b> o papel das turmas depois achar a sua fila e só depois do sinal você vai para aula.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

Conforme o Quadro 40, percebemos que a importância da família se faz presente também no Grupo do 8º ano/C20, sendo assunto presente em todas as propostas e, aqui, visualizamos, também nos diferentes anos pesquisados. Além disso, “amigos” é também uma palavra recorrente no *corpus* de estudo como um todo, apontando para o valor que os redatores dão para a amizade, conforme já mencionado anteriormente. Por fim, o verbo “achar” aparece com o sentido de

“encontrar” em todas as suas ocorrências, demonstrando que os alunos tendem a utilizá-lo sempre com esse sentido.

Como vimos, o perfil de vocabulário escrito dos diferentes grupos de redatores envolvidos na amostra é bastante próximo um do outro. Conforme discutido no Capítulo 2, os resultados de Biderman (1998a) e Nascimento e Isquardo (2013) apontaram para um possível núcleo de vocabulário, que seria utilizado em qualquer tipo de texto.

Apesar de idades, níveis de escolaridade, escolas e redes educacionais diferentes, o perfil lexical dos diferentes públicos analisados parece ser padrão: palavras gramaticais em maioria e semelhantes e palavras lexicais advindas principalmente dos textos de apoio, das propostas ou das temáticas de produção. Acreditamos que esse perfil segue o padrão da Língua Portuguesa como um todo, pois, conforme Antunes (2014, p. 32), enquanto as unidades lexicais “[...] têm índices de frequência variáveis, pois dependem do gênero e do tema em desenvolvimento [...], as unidades gramaticais [...] são de alta frequência e costumam ocorrer em todo tipo ou gênero de texto”.

## 6.2 RETOMADA DAS HIPÓTESES

As nossas três hipóteses são aqui retomadas, a fim de discutirmos sobre sua comprovação ou refutação.

**a) Os alunos dos dois principais grupos pesquisados apresentam diferenças significativas quanto ao perfil do vocabulário escrito no que se refere a/o(s):**

**l) repertório de formas gramaticais mais frequentes.**

Hipótese refutada. Em relação ao repertório de formas gramaticais mais frequentes utilizadas pelos estudantes, de maneira geral, houve proximidade entre o vocabulário escrito dos dois principais grupos pesquisados.

Especificamente em relação às palavras gramaticais mais utilizadas nos textos dos dois grupos, podemos dizer que é semelhante, como é possível observar nos quadros e tabelas do capítulo anterior. Nesses quadros e tabelas, é possível visualizar que os estudantes utilizam formas muito próximas, repetindo grande parte das palavras gramaticais mais utilizadas.

No experimento principal, entre as 50 primeiras palavras mais frequentes, cerca de 70% delas foram itens gramaticais, independente do grupo. Desses 70%, formas como “o”, “que”, “e”, “eu”, “de”, “no”, “um”, “do”, “ele”, “para”, “mais”, “com”, “se”, “muito”, “todo”, “por”, “em”, “bom”, “sem”, “quando”, “outro”, “mas”, “depois”, “isso” se repetiram entre os dois grupos. No Quadro 41, a seguir, estão trechos das redações dos dois grupos nas quais algumas dessas formas foram utilizadas:

Quadro 41 – Trechos de exemplo – palavras gramaticais em comum

<b>Escola Municipal – GRUPO A</b>	<b>Escola Estadual – GRUPO B</b>
<b>Eu</b> contaria tudo, desde de quando <b>eu</b> nasci até os dias de hoje.	No meu filme <b>eu</b> era uma criança muito pobre que sonhava ser jogador de futebol.
<b>Eu</b> faria com que tudo de ruim que acontece na minha vida se tornasse algo <b>bom</b> ou só diminuir o que me deixa triste.	A minha opinião é que o filme foi bem elaborado, divertido o filme é <b>bom</b> e facil de entender.
Se você for sair para a rua você tem risco de falecer, <b>mas</b> o Brasil dos meus sonhos não terá esse risco.	[...] quando ela viu as cartas não abriu, <b>mas</b> enviou a todas as pessoas marcadas nas cartas.

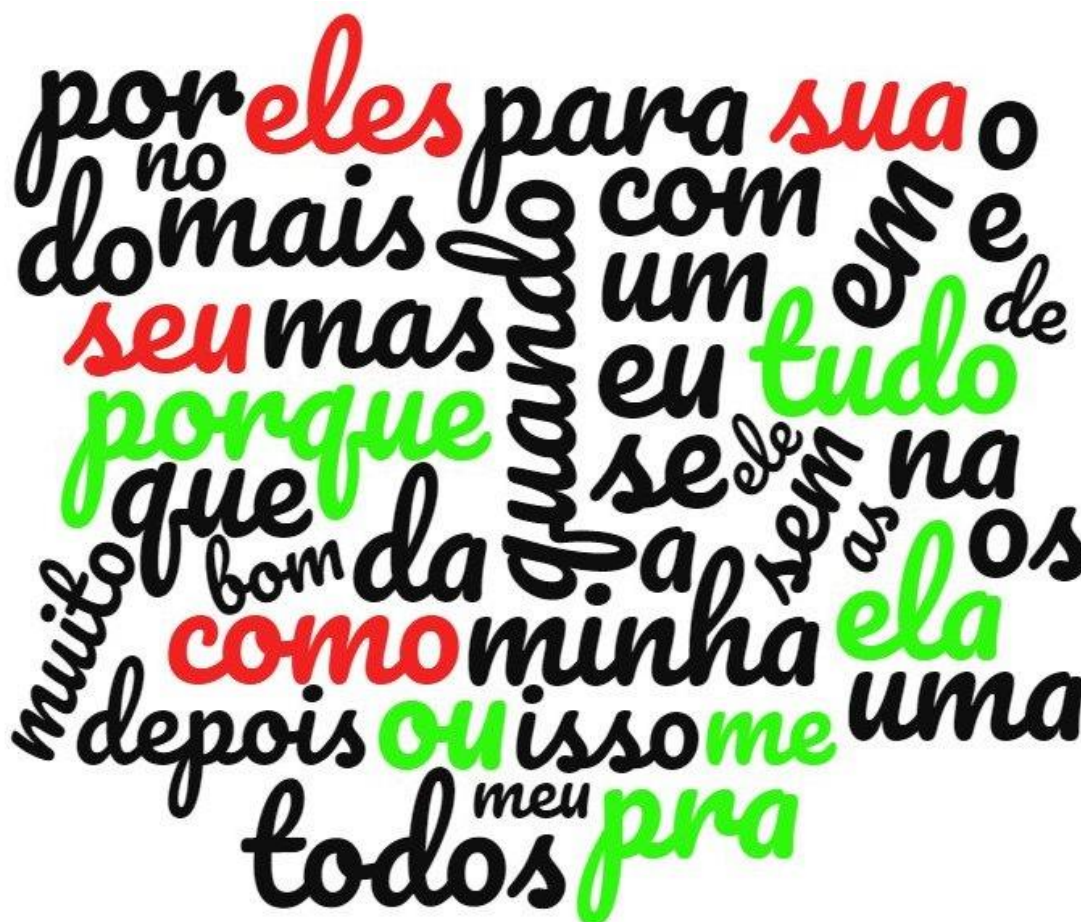
Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

A preocupação com o “eu” e o posicionamento dos estudantes em seus textos é algo recorrente e que também aproxima os grupos. Como podemos perceber no Quadro 41, ambos os grupos de redatores se colocam nos seus textos, apresentando suas opiniões. Isso também pode estar relacionado ao fato das propostas (ANEXO II) solicitarem e, de alguma forma, se prestarem a tais construções e posicionamentos. Ainda assim, percebemos, na leitura das redações, que os alunos têm o que dizer e dizem, expondo suas ideias e suas visões de mundo.

A coincidência de repertório gramatical entre as 50 palavras mais frequentes utilizadas nos textos dos dois principais grupos representou cerca de 85%. Isso fica visível na Figura 10:

Figura 10 – Nuvem de palavras gramaticais



Legenda: Verde – palavras gramaticais – Grupo A.  
 Vermelho – palavras gramaticais – Grupo B.  
 Preto – palavras gramaticais em comum nos dois Grupos.  
 Fonte: Elaboração própria.

A Figura 10 representa a nuvem de palavras gramaticais do nosso experimento principal, com o todo do *corpus* de estudo. Em vermelho e em verde estão as poucas palavras que aparecem somente num grupo. Em preto estão as palavras que aparecem em ambos os grupos pesquisados. Como podemos ver, elas são a grande maioria. Dessa forma, a quantidade de coincidências demonstrou que não há diferença significativa entre o repertório de formas gramaticais dos dois principais grupos pesquisados, refutando nossa hipótese.

Biderman (1998) cita a evolução das línguas românicas, do latim e de seus paradigmas e declinações e, em seguida, relata os estudos em outros idiomas (como o francês, o espanhol, o italiano e o romeno), que, juntamente com os estudos do português, apresentaram certas características em comum em relação ao vocabulário. Para a autora, “em todas essas línguas as altas frequências são

habitadas pelos vocábulos gramaticais e por um número reduzido de palavras lexicais de significado geral” (BIDERMAN, 1998, p. 163). Isso foi, de maneira geral, o que observamos ao analisarmos as redações escolares de nossos alunos.

## II) repertório de formas lexicais mais frequentes.

Hipótese refutada. No que tange ao repertório de palavras lexicais, os textos dos alunos dos dois principais grupos pesquisados também apresentaram semelhança no seu vocabulário escrito mais frequente.

Ambos os grupos têm preferência pelas palavras-tópico das propostas de produção, bem como pelas mais utilizadas nos textos de apoio, igualmente, nas redações das duas escolas, conforme já mencionado nas questões anteriores. Alguns exemplos são as palavras: “ser”, “ter”, “Brasil”, “querer”. Nos estudos de Biderman (1998), a lista de verbos mais frequentes inicia justamente por “ser” e “ter”, entre outros, com valores aspectuais e modais. A autora destaca que “Estes vinte verbos registram altíssima frequência por serem reiterados continuamente no texto. Dada a centralidade do verbo na articulação do discurso, é normal a enorme repetição dessas palavras” (BIDERMAN, 1998, p. 174).

Em relação às palavras lexicais que não tiveram como fonte os textos de apoio, também percebemos proximidade nas preferências dos dois grupos. Como exemplo, podemos citar a utilização frequente dos substantivos “pessoa”, “casa” e “dia”, como é possível visualizar nos trechos retirados dos dois grupos de texto, expostos no Quadro 42, a seguir:

Quadro 42 – Trechos de exemplo – palavras lexicais em comum

Escola Municipal – GRUPO A	Escola Estadual – GRUPO B
Com 16 anos me formei e comecei a trabalhar, com o meu primeiro salario ajudei minha avó a construir a <b>casa</b> dela e reformar a <b>casa</b> dela.	No meu filme eu era um militar que tinha uma <b>casa</b> muito boa e vivia em paz.
Voltar para a <b>escola</b> é reconhecer que ainda precisa aprender muitas coisas, e que por mais que aprenda muito, nunca se sabe tudo.	[...] eu também espero que eles consertem as nossas ruas e arrumar as <b>escolas</b> por que algumas <b>escolas</b> estão precisando de muitos reparos, e mandar um bom salario para os professores, e é isso que eu quero que mude

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

Conforme o Quadro 42, acima, percebemos que a moradia é algo significativo na vida dos alunos, que se preocupam com o estado e com o conforto desse ambiente. Além disso, a escola, que, muitas vezes, é a segunda (ou até mesmo a



primeira) casa desses jovens, também é assunto frequente nas suas produções. Muitos redatores demonstram sua preocupação com as condições desses lugares, com a falta de recursos e com o abandono do poder público, apontando diversas vezes as diferenças entre a escola pública e a escola privada e manifestando o desejo de melhorias.

Portanto, ainda que as propostas de produção tenham sido variadas, o repertório de formas lexicais mais utilizado se aproximou em ambos os grupos. Isso é o que visualizamos na nuvem de palavras lexicais a seguir:

Figura 11 – Nuvem de palavras lexicais



Legenda: Verde – palavras gramaticais – Grupo A.  
 Vermelho – palavras gramaticais – Grupo B.  
 Preto – palavras gramaticais em comum nos dois Grupos.  
 Fonte: Elaboração própria.

A partir da Figura 11, conseguimos visualizar que a coincidência entre os repertórios é expressiva. Apenas sete das dezesseis palavras lexicais presentes entre as 50 palavras mais frequentes de ambos os grupos não se repetiram. Dessas sete, três foram de uso específico do Grupo A e quatro do Grupo B. Com isso, nossa hipótese de diferenciação mais uma vez foi refutada.

### **III) variedade do vocabulário.**

Hipótese refutada. Quando avaliamos o repertório de vocabulário escrito dos alunos das duas redes, não percebemos diferenças consideráveis na sua variedade, tanto que suas riquezas lexicais, em percentual, foram sempre próximas uma da outra em todas as fases de nosso estudo.

À medida que fomos aumentando a quantidade de redações analisadas, a riqueza lexical, como já era esperado, diminuiu. No entanto, a diferença entre os dois grupos sempre foi pequena em todos os pequenos estudos. Afinal, de acordo com Biderman (1998), mesmo na língua escrita, em que, teoricamente, existem mais possibilidades e opções, o vocabulário realmente usado pelo indivíduo não é muito variado.

No experimento principal, com todo nosso *corpus* de estudo, foi que essa medida nos surpreendeu, pois, apesar de números diferentes de *tokens* e de *types*, o resultado, proporcionalmente, foi igual para as duas escolas: 16,5%. Convém lembrar que, conforme nossas referências explicitadas no Capítulo 2, uma variedade vocabular baixa, como é o caso desses grupos de redações escolares, não significa que o texto seja de má qualidade ou venha a receber avaliação mais baixa. Pelo contrário, nos estudos com exames de proficiência (EVERS, 2013) e com redações de vestibular (FINATTO; CREMONESE; AZEREDO, 2008; EVERS, 2018) percebemos que os textos que apresentaram as menores riquezas lexicais foram os mais bem avaliados.

Dessa forma, a riqueza lexical foi um quesito que, mais uma vez, aproximou os dois perfis de vocabulário escrito pesquisados, refutando também esta hipótese. Esses números demonstram interesses e rede de vocabulário semelhantes entre os dois grupos, apontando para um nível comum de desenvolvimento da linguagem, quando tomamos como base o aspecto lexical.

### **IV) especificidade(s) do vocabulário.**

Hipótese confirmada. Quando considerados os itens de frequência única, ou seja, as escolhas pessoais de cada grupo de alunos, percebemos a diferença no repertório vocabular dos estudantes.

Conforme já mencionado no Capítulo 2, nesse sentido estava o nosso interesse particular nessas palavras de frequência um. Por elas serem maioria no vocabulário de uma língua, quando realizada pesquisa léxico-estatística descritiva

com *corpora*, era possível que aí estivessem as diferenças de vocabulário que apostamos que existiriam entre os dois grupos de textos estudados. E os resultados mostraram que realmente há diferença entre esse grupo de itens referente ao vocabulário das duas escolas.

Para analisarmos o grupo de palavras de única ocorrência (*hapax*) dos dois principais grupos de textos dos estudantes pesquisados, optamos, conforme já explicado, por recortar as palavras iniciadas pela vogal A e pela consoante P. No início da pesquisa, nos primeiros estudos piloto, havia coincidência expressiva dessas palavras entre os resultados dos dois grupos, chegando a 67% de semelhança no estudo piloto II.

Foi nesse momento que percebemos a necessidade de lematizar as listas de palavras. Afinal, verificamos que muitos dos itens que ali estavam, só apareciam como itens de única ocorrência por encontrarem-se conjugados ou flexionados nas suas diferentes formas. As formas verbais “acaba”, “acabar” e “acabou”, “passa”, “passam”, “passar” e “passaria”, por exemplo, além de contabilizar itens diferentes nesse grupo *hapax*, ainda apareciam como elementos de única ocorrência, pois a ferramenta as identificou separadamente, cada palavra como uma forma.

Em função disso, a fim de tornar nossos dados mais acurados, resolvemos lematizar nossas listas. Assim, todas as flexões de uma mesma palavra foram agrupadas em sua forma básica ou canônica.

A partir do estudo piloto III, com a parceria dos cientistas da computação do NILC-USP, conforme já explicado neste trabalho, nossas listas de palavras foram automaticamente lematizadas. Esse processo, também já explicado, permitiu que enxergássemos melhor as preferências dos alunos, pois à medida que a camada *hapax* diminuiu, ele tornou-se também mais específico.

Os estudos seguintes, então, apresentaram diminuição considerável na coincidência de itens presentes no *hapax* desses dois grupos. A porcentagem de coincidência passou para apenas 10% e 5%, por exemplo, no estudo piloto III. Isso aconteceu porque palavras como “acontecer”, “aconteceria”, “aconteceriam”, “acontecido”, “aconteceram” e “aconteça”, que antes eram contadas separadamente, passaram a ser contabilizadas uma única vez como sua forma canônica, “acontecer”, por exemplo.

Além disso, formas que apareciam repetidas vezes no *hapax*, como o exemplo acima, do verbo “acontecer”, não pertenceram mais aos itens de única

ocorrência. Esses novos resultados encaminhavam, então, para a comprovação desta hipótese.

No experimento principal, apesar das quantidades de *hapax* dessas duas letras em cada grupo de textos serem semelhantes, quando analisadas as coincidências, obtivemos apenas 9% e 11% de palavras repetidas, respectivamente. Isso fica evidente nas Figuras 12 e 13, a seguir:

Figura 12 – *Hapax* em comum – Letra A



Legenda: Vermelho – *hapax* – Grupo A.

Azul – *hapax* – Grupo B.

Roxo – *hapax* comum nos dois Grupos.

Fonte: Elaboração própria.

Figura 13 – *Hapax* em comum – Letra P

Legenda: Vermelho – *hapax* – Grupo A.

Azul – *hapax* – Grupo B.

Roxo – *hapax* comum nos dois Grupos.

Fonte: Elaboração própria.

A partir das Figuras 12 e 13, conseguimos enxergar quão pequena é a coincidência de itens no *hapax* dos dois grupos. Os itens escritos nas cores azul e vermelha apresentam as palavras existentes apenas em cada um dos grupos. Vemos que são a maioria, seja nos itens que iniciam por A, seja nos itens que começam com P.

Como exemplo das diferenças observadas nos *hapax* iniciados pela vogal A, podemos citar “abc”, “adubar” e “alicerce”, nos textos da escola municipal. Nos textos da escola estadual, temos “abolir”, “amadurecimento” e “aspirante”. O Quadro 43, a seguir, mostra os usos dessas palavras nas redações escolares:

Quadro 43 – Exemplo de *hapax* específico de cada Grupo – vogal A

Escola Municipal – GRUPO A	Escola Estadual – GRUPO B
[...] destacaria também as minhas formaturas ( <b>ABC</b> , minha conclusão de término do ensino fundamental I e a minha formatura do curso de inglês.	Feministas devem ter mais voz de poder para buscar a igualdade e <b>abolir</b> o machismo.
Coisas para ajudar a poluir menos: separar todos os tipos de lixo, não jogar lixo no chão, reciclar, <b>adubar</b> , não pichar, etc...	Acho que o que eu destacaria na história seria seu crescimento pessoal, seu <b>amadurecimento</b> e sua nova visão da vingança, vendo que não era a melhor solução para seus conflitos pessoais.
Nossos governantes pensem mais na educação, segurança, e saúde do nosso Brasil pois a educação é o principal <b>alicerce</b> para o nosso futuro.	Fiquei dois anos no time de <b>aspirantes</b> ganhando muitos títulos até que eu subi para o time principal e fiquei por lá por mais três anos ganhando libertadores e etc,

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

No Quadro 43, é possível perceber, então, que, na especificidade, os alunos se expressam com diferentes itens de vocabulário. Escolhemos tais exemplos para demonstrar como, às vezes, os estudantes utilizam inclusive palavras diferenciadas, não frequente no uso dessa faixa etária. Além disso, em alguns trechos, é possível perceber a opinião e o senso crítico por parte desses redatores que, mesmo sendo jovens alunos de Ensino Fundamental, enxergam o que acontece a sua volta e se posicionam frente à realidade.

No Quadro 44, trazemos novos contextos de *hapax* iniciados pela consoante P:

Quadro 44 – Exemplo de *hapax* específico de cada Grupo – consoante P

Escola Municipal – GRUPO A	Escola Estadual – GRUPO B
Um Brasil com menos crimes com menos <b>pedofilia</b> [...]	Mas aqui na nossa já aconteceu de os professores dar aulas pros alunos e não ganhar o salário deles, eles até ganham mais <b>parcelado</b> .
O <b>protagonista</b> principal seria uma pessoa muito importante para mim.	Cada um tem sua <b>peculiaridades</b> e juntos estão sempre bagunçando e se divertindo muito.
E se o meu eu de hoje em dia olhasse para o de antigamente ia pensar: - <b>Porra</b> , mas que saudades daquele tempo.	Bom pulando para minha vida adulta, como falei eu quero servir, e lá dentro eu quero me formar para ser <b>perito</b> em corpos.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

Assim como nos exemplos anteriores, o Quadro 44 nos mostra usos singulares dessas palavras pelos estudantes. Além disso, seus contextos demonstram que os estudantes não só conhecem, como também sabem usar tais palavras. Mostram, ainda, que os alunos seguem se posicionando e expondo suas

opiniões. Dessa forma, foi nesse grupo de itens de frequência única que encontramos a principal diferença entre os perfis de vocabulário escrito pesquisados em nosso estudo.

Estão presentes nas Figuras 12 e 13 itens como “ponhar”, “apoucar”, “atendar”. Ao verificarmos seus contextos no *corpus*, concluímos que o lematizador automático que utilizamos estabeleceu tais formas como lema ou base dessas palavras. Isso não desqualifica o trabalho realizado, conforme já citado no Capítulo 5 deste trabalho.

A partir desses resultados, nossa hipótese de que os alunos dos dois principais grupos pesquisados apresentam diferenças quanto ao perfil do vocabulário escrito no que se refere às especificidades do vocabulário foi confirmada. Após lematizado o *hapax*, percebemos uma diferença expressiva entre esses itens nos dois grupos.

**b) Os alunos apresentam influência do vocabulário do(s) texto(s) de apoio nas suas redações.**

Hipótese confirmada. Conforme nossos estudos piloto, em que contrastamos o vocabulário mais frequente das redações dos alunos com o dos textos de apoio, podemos afirmar que os alunos são influenciados por esses materiais ao escreverem suas redações.

O primeiro e o segundo estudos piloto pesquisaram textos sobre a mesma proposta de produção textual: “A vida como filme”. Na nuvem de palavras a seguir, é possível visualizar as coincidências entre o repertório vocabular dos alunos e do texto de apoio:



Figura 14 – Nuvem de palavras – Texto de apoio / corpus de estudo



Legenda: Verde – palavras específicas do texto de apoio.  
 Vermelho – palavras semanticamente comuns entre os dois grupos.  
 Preto – palavras específicas do *corpus* de estudo.  
 Fonte: Elaboração própria.

Dessa forma, foi possível perceber que os alunos, ao escreverem as suas redações, utilizaram com frequência, além de muitas palavras gramaticais, também as seguintes palavras lexicais (entre as mais usadas na proposta ou no texto de apoio): os substantivos “filme”, “vida”, “coisa” e o verbo “ser”. Outros exemplos de itens usados pelos alunos que fazem parte do texto de apoio, mesmo não estando entre os mais frequentes, foram: “dia”, “amigos”, “foi”, “história”. O Quadro 45, a seguir, mostra os usos de algumas dessas palavras nas redações escolares frente aos usos no texto de apoio:



Quadro 45 – Exemplos – proposta 01

<b>GRUPO A – Escola Municipal</b>	<b>TEXTO DE APOIO – A vida como filme</b>	<b>GRUPO B – Escola Estadual</b>
Na verdade se minha <b>vida</b> fosse um <b>filme</b> , nunca mais ia voltar, ia me casar com um americano, ter filhos lindos e uma casa top.	O <b>filme</b> de nossa <b>vida seria</b> , independentemente de orçamento, uma obra-prima.	Pois minha <b>vida</b> é agitada, iria <b>ser</b> muito engraçado porque eu iria contar todos os meus “podres” da minha <b>vida</b> .
<b>Seria</b> um <b>filme</b> bem grande, com momentos tristes e felizes e os atores e atrizes seria Meus <b>amigos</b> e minha família.	Se pudéssemos fazer de nossa <b>vida</b> um <b>filme seria</b> , não tenhamos dúvida, um grande <b>filme</b> .	Se eu pudesse fazer um <b>filme</b> da minha <b>vida</b> , esse <b>filme</b> seria muito animado e também muito louco [...]
Meu <b>filme</b> iria <b>ser</b> de superação onde e minha família venceríamos todas dificuldades.	<b>Seríamos</b> os diretores do nosso <b>filme</b> . Com amplos e ilimitados poderes.	la <b>ser</b> o <b>filme</b> que contava minha adolescência, sobre as festinhas que eu fui, as minhas historias romanticas.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

No Quadro 45, acima, podemos perceber que, além das palavras que podem ser consideradas tema da proposta, as construções também são próximas do texto de apoio: “se minha vida fosse um filme”, “se eu pudesse fazer um filme da minha vida”, “o filme da minha vida”. Ainda assim, os alunos expressam seus gostos, suas preferências e seus sonhos ao colocarem o que gostariam no seu filme e como seria esse roteiro.

No terceiro estudo piloto, quando contrastamos os conjuntos de textos de cada proposta específica com o seu texto de apoio correspondente, percebemos que essa influência permanece. Além de muitas das palavras gramaticais serem semelhantes, também encontramos palavras lexicais presentes nas produções dos alunos e nos textos de apoio.

No estudo referente à proposta de número três, “Voltar de férias é”, encontramos as palavras “é”, “aula(s)” e “dia” entre as mais frequentes dos textos dos alunos e no texto de apoio. Outras palavras do texto de apoio que os alunos também utilizaram nas suas produções foram: “voltar”, “escola”, “professores”, “férias”, “colegas”. O Quadro 46, a seguir, traz exemplos de usos dessas palavras nas produções escolares e no texto de apoio:

Quadro 46 – Exemplos – proposta 03

GRUPO A – Escola Municipal	TEXTO DE APOIO – Voltar de férias é
<b>Volta</b> as aulas <b>é</b> retornar os estudos e aprendizados e recomeçar a ver as pessoas.	<b>Voltar</b> de férias <b>é</b> retomar o contato (ao vivo e a cores) com a galera do colégio.
<b>Voltar</b> as <b>aulas</b> <b>é</b> ter seu material e cuidar para que no final do ano tenha ainda	<b>É</b> retomar os papos da <b>aula</b> de história e as intermináveis discussões na <b>aula</b> de geografia.
O motivo bom é que não precisamos se preocupa com tema, acorda cedo e com <b>dia</b> de prova, o lado ruim é que temos que dormir cedo, acorda cedo, compra tudo novo e atura gente chata!	<b>É</b> aproveitar muito o único <b>dia</b> letivo sem provas, trabalhos ou professores de mau-humor.

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

No Quadro 46 percebemos as inquietações dos alunos na volta às aulas. Algumas delas, como “recomeçar a ver as pessoas” e “dia de prova” foram inspiradas nas ideias da autora do texto. Entretanto, outras, como “ter e cuidar do material” e “aturar gente chata” demonstram as preocupações e o universo desses jovens e adolescentes quando retornam às escolas no início do ano letivo.

Na análise referente à proposta de número 05, “O Brasil que eu quero”, as palavras coincidentes entre as mais frequentes foram: “é”, “Brasil” e “futuro”. Além dessas, encontramos outras palavras usadas pelos alunos que também estão no texto de apoio, como, por exemplo: “quero”, “escola”, “professores”, “país”. O Quadro 47, a seguir, mostra tais palavras, conforme foram usadas pelos alunos e no texto de apoio:

Quadro 47 – Exemplos – proposta 05

GRUPO B – Escola Estadual	TEXTO DE APOIO – Futuro
Bom as escolas públicas deveriam ter o ensino bom que nem a particular! E é esse <b>brasil</b> que eu quero para o nosso <b>futuro</b>	Isso me faz lembrar que o motivo desta gravação imaginária é dizer que <b>Brasil</b> eu quero para o <b>futuro</b> .
O <b>Brasil</b> que eu quero para o <b>futuro</b> , com mais pessoas humilde com mais igualdade, não esse <b>Brasil</b> sujo com preconceito.	Que tal um <b>Brasil</b> voltado para o <b>futuro</b> do <b>futuro</b> , em que as crianças recebam proteção (e limites) em casa, ensino qualificado na escola e, na vida, o desafio permanente de construir um país melhor?

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

Ao observarmos o Quadro 47, percebemos que os alunos têm opinião e se expressam ao escreverem seus textos. Eles enxergam a diferença da sua escola para uma escola particular, seja na falta de estrutura, de material ou de professores. E isso aparece nos seus textos, pois é algo importante para eles. Também buscam a

igualdade nesse Brasil que desejam para o futuro, um país com pessoas mais educadas e humildes.

Nossa hipótese B foi confirmada, portanto, uma vez que há influência dos textos de apoio no vocabulário escrito usado pelos alunos nas suas redações. Conforme nossos exemplos, os alunos utilizam palavras dos textos de apoio nas suas produções, o que é esperado, na medida em que é natural que a temática da proposta seja utilizada pelos alunos com frequência ao redigirem seus textos. Apesar disso, percebemos também a manifestação dos redatores, que não se preocupam apenas em completar as linhas solicitadas, mas que se colocam nas redações, expondo suas opiniões e seus desejos, demonstrando seus principais interesses. Ainda que com escritas e construções próximas da oralidade, percebemos que os estudantes têm senso crítico dos fatos e da realidade a sua volta. Isso tende a acontecer em textos como os que compõem o *corpus* de estudo, pois o caráter mais opinativo das propostas (ANEXO II) se presta a tais construções.

Tais resultados corroboram com dados das pesquisas de Nascimento e Isquierdo (2003) que, ao compararem um *corpus* de redações de vestibular ao *Dicionário de frequências* de Biderman, concluíram que mais de 85% das palavras mais frequentes coincidem. Tal fato, bem como os nossos resultados, vão ao encontro dos achados de Biderman sobre algumas palavras serem consideradas “multiuso” e estarem presentes em todo e qualquer tipo de texto, independente do seu gênero e/ou temática. Nosso estudo usou como *corpus* principal textos do gênero redação escolar, e, como *corpus* de contraste, textos de apoio às propostas, do gênero crônica. Ainda assim, com gêneros diferenciados, houve proximidade no vocabulário escrito, colaborando com a validade do princípio das palavras “multiuso”.

**c) O vocabulário escrito dos alunos é próximo de um padrão de português popular escrito, representado pelo CorPop.**

Hipótese confirmada. Ao compararmos as listas de palavras lematizadas do CorPop e do *corpus* de estudo – com todas as 154 redações variadas dos alunos – percebemos que 67,4% das palavras do *corpus* de estudo estão presentes também no CorPop. Ressaltamos que essa proximidade foi verificada em termos de distribuição e não de frequência entre os *corpora*.

É importante lembrar que a lista lematizada do CorPop passou por um processo de “limpeza”, que incluiu a exclusão de muitas das palavras gramaticais

presente nesse *corpus*, além de outros itens como dias da semana e nomes de cidades, de países e de localidades. Então, essa quantidade de palavras coincidentes entre os dois *corpora* provavelmente aumentaria caso na lista lematizada do CorPop estivessem esses outros itens.

Podemos citar como exemplo de coincidências algumas palavras como: “achar”, “bairro”, “capacidade”, “dependente” e “eliminar”. Em quase todas as letras do alfabeto foram encontrados itens do CorPop em nosso *corpus* de estudo, com exceção somente das letras “K”, “W”, “Y”<sup>29</sup> e “Z”, o que não nos surpreende, pois são letras com pouca produtividade em nosso idioma. A seguir, o Quadro 48 exemplifica os usos de duas palavras: “achar” e “pobre”.

Quadro 48 – Exemplos de uso – CorPop/*corpus* de estudo

<b>CorPop</b>	<b>Corpus de estudo</b>
<ul style="list-style-type: none"> <li>- Também é difícil <b>achar</b> um lugar onde possa fazer fogo e cozinhar.</li> <li>- Não consegui <b>achar</b> a cura para o caso.</li> <li>- Faça o que <b>achar</b> melhor, só achei que gostaria de considerar isso.</li> <li>- Ele vai <b>achar</b> você bonita!</li> <li>- É preconceito <b>achar</b> que quem toca violão é um desclassificado.</li> </ul>	<ul style="list-style-type: none"> <li>- Micky <b>acha</b> o um mapa do tesouro, após <b>achar</b> o mapa Micky e seus saiem para procurar o tesouro</li> <li>- começam a vasculhar o local ate <b>achar</b> uma passagem secreta debaixo de uma lareira</li> <li>- daí eles conseguiram <b>achar</b> o tesouro num navio.</li> <li>- você tem que chegar na escola, passar pela multidão depois <b>achar</b> o papel das turmas depois achar a sua fila</li> <li>- passar pela multidão depois achar o papel das turmas depois <b>achar</b> a sua fila e só depois do sinal você vai</li> </ul>
<ul style="list-style-type: none"> <li>- O cachorro pode passear, mas o <b>pobre</b> não.</li> <li>- <b>Pobre</b> não tem direito a segurança, só rico que tem.</li> <li>- Filho de <b>pobre</b> é assim.</li> <li>- até o comendador desistir de violentar a <b>pobre</b> escrava.</li> <li>- <b>Pobre</b> de minha irmã!</li> <li>- <b>Pobre</b> não pode dançar, não pode cantar.</li> </ul>	<ul style="list-style-type: none"> <li>- No meu filme eu era uma criança muito <b>pobre</b> que sonhava ser jogador de futebol.</li> <li>- vim de família <b>pobre</b> area rural tinha que ir de ape para o colegio</li> <li>- Mesmo Michael sendo negro e <b>pobre</b> ele conseguiu alcançar o seu objetivo que era ser jogador</li> <li>- somos todos iguais, não enporta se sou <b>Pobre</b> ou Rica, pequena ou grande</li> <li>- vou dar varios exemplos: um negro crescido em periferia <b>pobre</b> se tornar rico e que todos admiram</li> <li>- Eu mesmo sabendo que o Brasil é um país <b>pobre</b> meu sonho é viver em um país “rico”</li> </ul>

Legenda: Negrito – palavras analisadas.

Fonte: Elaboração própria.

<sup>29</sup> A partir do Novo Acordo Ortográfico (2009), o alfabeto português passa a ter 26 letras (a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z). Foram oficialmente incorporadas três (k, w e y), que podem ser usadas em casos especiais, como para o registro de nomes próprios ou alguns empréstimos em obras lexicográficas, por exemplo.

Conforme o Quadro 48, acima, percebemos que os usos, no CorPop, têm sentidos diferentes para mesma palavra, enquanto no *corpus* de estudo o sentido é o mesmo em todas as ocorrências. O primeiro exemplo, “achar”, tem o sentido de “encontrar” ou de “pensar” no *corpus* de referência, enquanto no *corpus* de estudo o único sentido é o de “encontrar”. O mesmo acontece no segundo exemplo, em que “pobre”, no CorPop, tem o sentido de “pessoa com pouco/sem bens ou dinheiro”, “coitado(a)”, e é usado como substantivo algumas vezes, substituindo “a pessoa/o indivíduo/o ser humano pobre”. Já no *corpus* de estudo, o sentido e o uso são somente do adjetivo que caracteriza seja a criança, a família, o Michael, o autor do texto (eu), a periferia ou o Brasil. Os trechos de exemplo apontam para um uso mais concreto das palavras pelos alunos do Ensino Fundamental, que preferem utilizá-las com único sentido, o primeiro e mais comum dos sentidos possíveis para cada palavra. No *corpus* de referência, que reúne maior quantidade de textos, percebemos que o uso é diverso, explorando possibilidades de construções e de sentidos dessas palavras.

Em relação a tais usos diferenciados, salientamos que os textos que formam o CorPop têm como referência de leitor “[...] indivíduos entre 15 e 64 anos, das classes B, C e D/E, em níveis elementar e intermediário de letramento” (PASQUALINI, 2018, p. 67). Segundo dados do INAF (2018),<sup>30</sup> esses níveis são compostos, em maioria, por indivíduos que ingressaram ou concluíram o Ensino Médio (INAF, 2018, p. 11, Tabela 3a). Dessa forma, a coincidência de cerca de 70% entre os itens dos dois *corpora* se justifica, uma vez que o *corpus* de estudo é formado por redações de alunos de Ensino Fundamental, ainda em processo de letramento. Além disso, é importante lembrar que o *corpus* de referência passou por um processo de ‘limpeza’, conforme já explicado. Nesse processo, alguns itens, sobretudo palavras gramaticais, foram retirados do *corpus*. Então, essa porcentagem de coincidência provavelmente seria maior caso tais itens integrassem o *corpus* de referência.

De acordo com Correia (2020), a competência lexical varia de indivíduo para indivíduo, pois se desenvolve ao longo da vida. Em função disso, as pessoas não dominam os mesmos itens, uma vez que isso varia ao longo da idade e das experiências. Segundo a autora, tal competência depende de fatores externos como:

---

<sup>30</sup> Disponível em: <[https://acaoeducativa.org.br/2018/08/Inaf2018\\_Relat%C3%B3rio-Resultados-Preliminares\\_v08Ago2018.pdf](https://acaoeducativa.org.br/2018/08/Inaf2018_Relat%C3%B3rio-Resultados-Preliminares_v08Ago2018.pdf)>. Acesso em: 20 ago. 2020.

escolarização, meio sociocultural, história de vida, entre outros. Então, nossos resultados parecem apontar para um bom indicativo da qualidade do trabalho que está sendo feito na escola pública, pois os quase 70% de coincidência demonstram que o nível e o amadurecimento do repertório de vocabulário escrito desses jovens estão em processo, em construção.

Outro ponto que merece considerações é o fato de não termos um levantamento específico do vocabulário fundamental escrito do português, conforme utilizado por estudantes do Ensino Fundamental, de acordo com as suas frequências de uso e distribuição em suas produções escritas. Além disso, ainda que seja possível argumentar que qualquer um dos três *corpora* de referência utilizados neste trabalho podem nos fornecer uma lista de vocabulário comum mais frequente, além de possibilitarem um conjunto vocabular de intersecção entre eles, não há como afirmar que essa conexão corresponda a um vocabulário fundamental da Língua Portuguesa. Não obstante, permanece o questionamento sobre quais seriam os pontos de corte em termos de frequência para o estabelecimento do que seria fundamental ou não em um *corpus* de caráter geral que não incluía redações escolares.

Infelizmente, nossas investigações estão longe dos estudos com a língua inglesa, por exemplo. O que temos publicado são os experimentos de Biderman sobre a frequência dos verbos, pois seu dicionário completo não chegou a ser publicado. Nesses estudos, uma de suas descobertas é que:

Cerca de 42% do total das ocorrências do corpus é constituído por pouco mais de mil palavras, as mais freqüentes da língua (ver Gráfico 1). Este gráfico atesta o seguinte: 80% de qualquer texto do português é constituído por estas mil palavras, que são reiteradas continuamente. (BIDERMAN, 1998a, p. 166).

Então, segundo Biderman (1998a), qualquer tipo de texto terá essas palavras, que funcionariam como uma espécie de “vocabulário fundamental” da língua. Com relação ao português, isso já tinha sido comprovado por Duncan (1972), com um *corpus* menor, mas com resultados muito semelhantes. Além disso, resultados parecidos também foram obtidos em pesquisas com outras línguas românicas. Em função disso, Biderman (1998a, p. 167) acredita que “[...] podemos extrapolar estes resultados e formular a hipótese de que o mesmo deve ocorrer em qualquer língua”.

No português brasileiro, essas palavras seriam as que chamamos, nesta pesquisa, de palavras gramaticais e algumas palavras lexicais. Conforme a autora:

As palavras gramaticais e certo grupo de verbos são as classes mais estáveis da língua. São palavras multiuso que aparecem em qualquer texto, independentemente de seu conteúdo temático. Donde a distribuição uniforme dessas palavras entre os vários gêneros escrutinados. De fato, pode-se constatar que elas são igualmente freqüentes em todos os cinco tipos de literatura analisados, seja qual for o gênero ou o tema tratado. O mesmo pode ser dito, com algumas ressalvas, de um grupo de adjetivos e substantivos de significado muito geral. (BIDERMAN, 1998a, p. 168).

Biderman (1998a, p. 168) acredita que “estas palavras constituem o núcleo do vocabulário do português e podem ser consideradas como essenciais para a comunicação neste idioma”. Nesse sentido, acreditamos que os resultados de pesquisa, quando comparados ao principal *corpus* de referência, vão ao encontro dos resultados de Biderman, na medida em que apontam para uma espécie de “núcleo” de vocabulário, independente do gênero textual e da temática.

Em função disso, acreditamos que, com este estudo, contribuímos para pensar e refletir sobre a nossa língua. O vocabulário de fato utilizado pelos alunos de Ensino Fundamental pode e deve auxiliar nas atividades de ensino de língua, de exploração e de aquisição de vocabulário. Foi a partir disso que idealizamos as atividades que serão expostas no Capítulo 7.







### 6.3 SÍNTESE

Em linhas gerais, as nossas questões de pesquisa se orientaram em direção às possíveis diferenças de vocabulário escrito manifestado nas produções textuais dos estudantes das duas realidades envolvidas. A expectativa contida nas questões era de heterogeneidade. Entretanto, as respostas iluminaram, justamente, as semelhanças, apontando para a homogeneidade entre os grupos de textos.

Frente a isso, vale registrar que as inquietações que nos levaram a realizar esta pesquisa foram justamente as diferenças percebidas subjetivamente entre os dois públicos, tanto por nós, quanto por outros colegas professores das duas realidades. Tais distinções não se apresentaram na prática, nos resultados do estudo científico, demonstrando que, do ponto de vista léxico-estatístico, há proximidade entre esses dois públicos.

A Figura 15, a seguir, apresenta um resumo dos nossos achados acerca de cada hipótese:

Figura 15 – Síntese das hipóteses

<i>Retomada das Hipóteses</i>	
Hipótese	Resultado
a) Os alunos dos dois principais grupos pesquisados apresentam diferenças quanto ao perfil do vocabulário escrito no que se refere a(s): I) repertório de formas gramaticais mais frequentes;	
II) repertório de formas lexicais mais frequentes;	
III) variedade do vocabulário;	
IV) especificidade(s) do vocabulário.	
b) Os alunos apresentam influência do vocabulário do(s) texto(s) de apoio nas suas redações.	
c) O vocabulário escrito dos alunos é próximo de um padrão de português popular escrito, representado pelo CorPop.	

Fonte: Elaboração própria.

Conforme a Figura 15, os perfis de vocabulário escrito analisados se diferenciam somente na especificidade, nas suas escolhas particulares. Além disso, apesar de apresentarem influência dos textos de apoio, os alunos redatores têm opinião e se colocam nos seus textos. Por fim, podemos dizer que a aquisição de vocabulário parece estar em uma escala ascendente em relação ao que vemos no CorPop.

Consideramos, inspiradas em Correia (2020), que a competência lexical é fator de discriminação ou de promoção social e, por isso, é facilitadora da aprendizagem e do sucesso educativo. Ou seja, quem tem maior competência lexical é quem é mais exposto e tem mais acesso. Nesse sentido, acreditamos no reconhecimento do repertório lexical dos nossos alunos, bem como no seu alargamento, por meio de atividades e da exploração constante dessa temática em sala de aula, como forma de promover esse acesso e essa exposição.

Em função disso, no capítulo seguinte deste trabalho, apresentamos sugestões de atividades para uso em sala de aula de língua materna do Ensino



Fundamental. De acordo com Assunção (2015, p. 07), “[...] constatamos que há necessidade de um trabalho didático-pedagógico mais efetivo com o léxico no Ensino Fundamental”. Dessa forma, buscamos contribuir com a lacuna de estudos nessa área.

Essa proposta está embasada em nossa experiência docente e nos dados que obtivemos do nosso *corpus* de estudo, bem como nos resultados dos nossos experimentos, pois, de acordo com Silva (2016, p. 123):

Além de quantificar é necessário observar esses padrões linguísticos e analisar o mecanismo de funcionamento da língua em questão, tornando o planejamento para ensino ou aprendizado mais eficiente, uma vez que, por meio da pesquisa, conseguimos compreender o sistema linguístico do idioma.

Dessa forma, esperamos, com tais propostas de ensino, auxiliar na ampliação das competências escrita e leitora dos alunos das escolas públicas, por meio do alargamento do seu repertório lexical. Afinal, é importante também que o professor saiba explorar, didaticamente, o tema do vocabulário/léxico com os seus alunos dos diversos níveis escolares.

## **7 SUGESTÕES DE ATIVIDADES DE ENSINO DE VOCABULÁRIO**

Por meio das comparações e dos contrastes realizados, conseguimos descrever uma parcela do vocabulário escrito dos alunos de Ensino Fundamental das redes públicas, além de caracterizá-lo de acordo com a escola/rede e ano/ciclo. Dessa forma, com a devida preservação das grafias originais e de eventuais erros, organizamos também uma pequena base de dados.

Esses resultados, além de embasar as respostas das questões de pesquisa e a verificação das hipóteses, tornaram possível o desenho de sugestões de atividades de ensino de vocabulário. Como docentes de língua materna e pesquisadoras na área de léxico, acreditamos que esse estudo continuará agregando valor à nossa prática em sala de aula. Queremos, antes de tudo, poder qualificar nosso trabalho, e compartilhar, então, essa experiência com nossos colegas docentes, na forma dessas sugestões.

Acreditamos que as aulas de língua materna são espaços privilegiados para os estudos de vocabulário e, sobretudo, para o alargamento do repertório vocabular dos estudantes. Consideramos também que nosso estudo sobre padrões de vocabulário escrito de estudantes do Ensino Fundamental de escolas públicas é capaz de contribuir para isso.

Então, conforme um dos nossos objetivos, idealizamos atividades de ensino de vocabulário praticáveis em sala de aula, mesmo sem o auxílio de ferramentas computacionais e/ou internet (considerando a realidade da escola pública brasileira atual). Essas atividades são apresentadas neste capítulo e podem auxiliar o professor em sala de aula, contribuindo para a expansão do vocabulário dos seus alunos. Pensamos que, quanto maior o espectro vocabular do indivíduo, maior também será sua capacidade de leitura, de interpretação e de compreensão, conforme já abordado nos capítulos iniciais e ao longo de todo este estudo. Dessa forma, acreditamos que seja possível contribuir com a proficiência leitora desses jovens em formação escolar.

A ideia é que as atividades que sugerimos aqui sejam difundidas para os colegas, professores de língua materna, para que sejam realizadas com seus alunos em sala de aula. Sabemos que uma hipótese sobre alargamento do repertório vocabular dos alunos, a partir de atividades como essas, só pode ser validada após

a aplicação dos exercícios em sala de aula e da testagem. Coloca-se, assim, uma ideia para estudo futuro.

A despeito disso, acreditamos que, com tais sugestões de atividades, estamos contribuindo não só com a nossa prática, mas também com a de todos os colegas da área. Ao colaborarmos com a proficiência leitora dos estudantes, estaremos também auxiliando todas as outras disciplinas que também notam e questionam as dificuldades de compreensão dos nossos alunos, independente da área do conhecimento.

Nossas sugestões de atividades são apresentadas de duas formas diferentes. Primeiro, temos sugestão de uma unidade de ensino completa, com exercícios de leitura, de interpretação, de produção, de reescrita e de aspectos vocabulares. Na segunda parte estão propostas de atividades independentes, que podem ser efetuadas intercaladas com outros conteúdos e temas, não sendo necessário seguir uma ordem ou realizar todo o conjunto proposto.

## 7.1 PROPOSTA DE UNIDADE DE ENSINO DE VOCABULÁRIO

A unidade de ensino segue os princípios dos Referenciais Curriculares (RCs), procurando acompanhar seus passos, independente do conteúdo e do texto que serão trabalhados. Inicia com uma sondagem sobre o hábito de leitura dos alunos e o que eles leem. Após, há uma preparação para leitura por meio de figuras e discussões sobre o título do texto. Em seguida, idealiza a leitura silenciosa e individual do aluno, para que seja possível um primeiro contato com o texto e, na sequência, uma leitura em voz alta pelo grupo, para que seja exercitada também a oralidade. Continua com interpretação e estudo do texto e, ainda, uma produção textual.

Como nosso estudo aborda a temática do vocabulário escrito dos alunos, esta unidade de ensino se encaixa com o trecho em que os RCs (2009) tratam de “linguagem”, abordando essa temática, incluindo exercícios aplicados que utilizam como referência as próprias produções textuais dos alunos. Em função disso, pensamos num trabalho de estatística lexical (apoiado em BIDERMAN, 1978; 1998) com os próprios textos dos alunos, em que eles repensariam, a partir do vocabulário deles mesmos, por exemplo, quais palavras foram usadas em determinado contexto e o porquê, quais se repetem menos, quais se repetem mais, quais sinônimos

poderiam substituí-las e quais palavras não conhecem. Além disso, conforme preveem os RCs, há proposta de reescrita das produções textuais após exercícios aplicados de estatística vocabular e estudo pessoal das redações pelos próprios alunos redatores. Acreditamos que, por meio de tais atividades, os alunos refletirão acerca da sua produção escrita, do seu vocabulário e do seu repertório lexical. Essa reflexão, imaginamos, os permitirá avanços nas suas reescritas, que serão pontuados e divididos com o grupo ao final da unidade de ensino.

Os exercícios aplicados de estatística lexical são guiados pelos princípios teórico-metodológicos da Linguística de Corpus (BERBER SARDINHA, 2004), entretanto, os termos e a metodologia utilizados são adaptados ao contexto escolar de uma aula de Língua Portuguesa, sendo acessíveis aos professores e alunos do Ensino Fundamental das redes públicas.

Idealizamos como público-alvo das atividades os alunos do 8º ano do Ensino Fundamental regular de escolas públicas, pois foi principalmente a partir de produções escritas de alunos desse ano/ciclo que foi construído o *corpus* principal desta pesquisa: as redações escolares.

Acreditamos que um trabalho cuidadoso com a leitura e a escrita e, a partir delas, com léxico, é muito importante em todo o processo de ensino-aprendizagem. Portanto, nada impede que tais atividades sejam realizadas com outros anos/ciclos, desde que o objetivo seja mantido: o alargamento do repertório vocabular do aluno. Acreditamos que um vocabulário variado e extenso pode contribuir significativa e positivamente para o desenvolvimento da proficiência leitora e interpretativa do jovem estudante de Ensino Fundamental. Afinal, quanto mais diversificado for o seu repertório vocabular, maior também será sua facilidade de compreensão de frases, parágrafos, textos, contextos e, assim, das atividades e tarefas que lhe forem solicitadas, não só em sala de aula, mas na sua vida como um todo.

### **7.1.1 Unidade de Ensino de Vocabulário**

**Público-Alvo:** alunos de 8º ano

**Tempo estimado:** A unidade de ensino está separada por partes, na qual se imagina o cenário de 2 períodos (entre 90min e 100min) de aula de Língua Portuguesa para cada parte.

Obs.: O total de semanas e de períodos deve variar conforme a carga horária de períodos de Língua Portuguesa da escola e conforme o tempo estipulado para cada período.

**Avaliação:** Contínua, em que o aluno é avaliado principalmente conforme seu interesse, sua participação e seu engajamento nas atividades. Serão entregues 3 atividades, como forma de documentar essa avaliação: uma produção textual (escrita e reescrita), um relatório estatístico e uma autoavaliação.

**Parte 1:** Apresentação da unidade de ensino. Preparação para a leitura. Apresentação do texto: “A vida como filme” (crônica de Moacyr Scliar, publicada no jornal Zero Hora) (ANEXO III). Leitura (individual e em grupo). Interpretação.

➤ Apresentação da unidade de ensino: a unidade de ensino é iniciada com uma breve explicação da proposta, das atividades, do que será trabalhado e dos objetivos do projeto. Além disso, ressalta-se a participação dos estudantes como muito importante em todos os momentos. Em função disso, é salientado que participem sempre que acharem necessário e, principalmente, sempre que forem solicitados, pois a opinião de cada um é importante e sempre bem-vinda. Depois, é feita a chamada, para que seja possível a apresentação dos alunos.

➤ Preparação para a leitura: para iniciar, é feita uma sondagem sobre o hábito da leitura e o hábito de leitura de jornal dos alunos, de seus pais e familiares. Eles comentam o que preferem/gostam de ler (revistas, quadrinhos, literatura, *blogs*, *Instagram*, *Twitter*) e, no caso dos jornais, se têm conhecimento/hábito, qual(is) preferem, além de suas seções preferidas e justificativas. Depois, é abordada a questão dos jornais virtuais, do acesso às notícias e conteúdos pela internet, das *fake news* e da credibilidade dos meios eletrônicos.

➤ Apresentação do texto “A vida como filme”, de Moacyr Scliar: discussão sobre o título “A vida como filme” (a partir dele, do que os alunos acham que trata o texto) e sobre figuras coladas no quadro (ANEXO VI).

➤ Leitura: individual (cada aluno lê silenciosamente, para que seja possível um primeiro contato) e em grupo (cada aluno lê um trecho do texto para a turma,

exercitando, assim, a oralidade). Na sequência da leitura, algumas questões são discutidas oralmente, como os gostos pessoais de filmes, atores e atrizes, além de como fariam um filme, caso fossem roteiristas do próprio enredo.

➤ **Interpretação:** perguntas sobre o texto são colocadas no quadro (ANEXO VII), os alunos têm um tempo para responder/pensar sobre as questões que, depois, são discutidas no grande grupo. Após cada aluno responder com suas interpretações e opiniões, são levantadas, no grande grupo, as possibilidades de interpretação, além de algumas questões que vão surgindo ao longo de cada discussão.

**Parte 2:** Retomada do texto, do assunto e das questões de interpretação. Características do texto do tipo crônica. Proposta de produção textual.

➤ **Retomada do texto:** retomada oral do assunto do texto, das questões de interpretação e das discussões da aula anterior.

➤ **Características do texto do tipo crônica:** após a retomada do assunto do texto trabalhado e discutido na aula anterior, são apresentadas características deste tipo de texto, sempre com exemplos e trechos do texto já trabalhado. Os alunos percebem que a crônica é um texto sobre algo do cotidiano, produzido para ser publicado na imprensa, seja nas páginas de um jornal ou de uma revista, ou, atualmente, na internet. Comentam que nela há normalmente uma crítica, seja direta ou indireta. Enxergam, também, que assim como o repórter, o cronista se inspira nos acontecimentos diários, dando um toque pessoal ao texto. Na maioria dos casos, é curta e narrada em primeira pessoa, ou seja, o próprio escritor está "dialogando" com o leitor. A linguagem é simples, espontânea, situada entre a oralidade e a linguagem literária.

➤ **Proposta de produção textual:** após apresentadas as características de uma crônica, a partir de exemplificações retiradas do texto de Scliar, os educandos são instigados a escreverem suas produções, baseados nas leituras e nas interpretações realizadas. A proposta é de 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

(ANEXO VIII). As produções são entregues e avaliadas conforme combinado com os alunos.

**Parte 3:** Devolução e retomada das produções textuais. Apresentação da estatística lexical, da noção de palavra (utilizada em pesquisas estatísticas) e de padrões de frequência de palavras e de vocabulário. Exercícios aplicados.

- Devolução e retomada das produções textuais: as produções entregues pelos alunos são corrigidas e devolvida aos redatores, com sugestões de mudanças e melhorias (quando for o caso) e comentários individuais sobre o texto. Elas são usadas como *corpus* das atividades de estatística lexical realizadas na sequência da unidade de ensino.
- Apresentação da estatística lexical, da noção de palavra (utilizada em pesquisas estatísticas) e de padrões de frequência de palavras e de vocabulário: os alunos são apresentados à estatística lexical, por meio de exemplos com frases/textos, de pesquisas realizadas com padrões de vocabulário, de *corpora* existentes (ANEXO VI) e de exercícios. A partir dos exemplos, será possível perceber que a noção de palavra utilizada em estatística lexical é a de palavra gráfica, que, conforme já vimos neste trabalho (Capítulo 2) e de acordo com Bisognin (2009, p. 25) é “(...) toda unidade lingüística mínima que pode constituir significado, delimitada na escrita por dois espaços em branco e/ou sinal de pontuação”. Os alunos também familiarizar-se-ão com outros termos da área, como quantidade de palavras, repetições de palavras, frequências de palavras. Isso os auxiliará nos exercícios que serão aplicados.
- Exercícios aplicados: após a familiarização com a estatística lexical, os alunos são orientados a realizarem alguns exercícios aplicados, usando como *corpus* a sua produção textual que foi devolvida e preenchendo a avaliação, que consiste num relatório dos resultados encontrados nesses exercícios. O primeiro exercício consiste em contar o número de palavras (usando o critério de palavra gráfica, separada por espaços em branco) totais do seu texto. Em seguida, o segundo exercício consiste em anotar todas as palavras do seu texto, uma abaixo da outra, somando quando se repetem ao longo de sua produção. O terceiro exercício

consiste em organizar essas palavras em ordem alfabética. O quarto, em organizá-las por ordem decrescente de frequência. Todos esses dados e listas são anotados no relatório (ANEXO IX), que é entregue juntamente com a produção textual de cada aluno.

**Parte 4:** Retomada e continuidade dos exercícios estatísticos. Variedade vocabular.

➤ Retomada e continuidade dos exercícios estatísticos: os relatórios estatísticos entregues são devolvidos aos alunos, bem como suas produções textuais. Os alunos são orientados a selecionarem as palavras lexicais (aquelas com conteúdo nocional mais específico, associadas a fatos, processos, fenômenos) mais frequentes no seu texto e a buscarem os contextos em que aparecem. Após, são instigados a buscarem sinônimos, expressões ou construções pelas quais podem ser trocadas tais palavras (o uso de dicionários e/ou internet fica a critério do professor neste momento). Essas sugestões são anotadas a fim de enriquecer a reescrita da redação de cada aluno, aumentando seu repertório vocabular e estimulando a variação lexical na escrita.

➤ Variedade vocabular: a partir do trabalho com os contextos de uma mesma palavra e da busca por diferentes itens para serem usados com o mesmo sentido, a variedade vocabular é apresentada para os alunos. Como essa noção introduz a noção de proporção e regra de três, que faz parte dos conteúdos de Matemática, sugere-se trabalho interdisciplinar com essa disciplina, de maneira que os professores trabalhem juntos. Assim, a partir da interdisciplinaridade, os alunos são ensinados a realizarem a conta (ANEXO X) que leva ao resultado percentual de riqueza lexical de seu texto. Além disso, são mostrados exemplos de diferentes textos, com diferentes riquezas, como a crônica de Scliar, um texto científico, uma lista de supermercado e uma redação escolar (ANEXO XI). A ideia é que os alunos percebam o quanto a variedade vocabular tem influência nos textos e como pode auxiliar ou atrapalhar, dependendo do intuito do texto. Ainda com auxílio da disciplina de Matemática, é mostrado aos alunos como essas medidas estatísticas podem ser representadas por um gráfico (ANEXO XII). Eles aprendem a construir um gráfico de frequência de palavras, em que eixo Y corresponde ao número de repetições e o eixo X à quantidade total de palavras. Dessa maneira, é possível



perceber as palavras que mais se repetem num texto (que, geralmente, são as mais repetidas na língua portuguesa como um todo e são palavras gramaticais), pois estarão na ponta de cima do gráfico, até a primeira palavra lexical, que demonstra o tópico do texto e está próxima das palavras-tema do texto. A maior camada do gráfico é chamada de “cauda” e lá estão as especificidades de um texto, por isso tende a ser a camada mais extensa, quando analisadas as frequências de palavras. O último dos exercícios estatísticos aplicados é justamente a construção do gráfico correspondente ao texto de cada aluno. Os dados levantados neste momento também são documentados no relatório de estatística lexical, que novamente é entregue.

**Parte 5:** Retomada dos exercícios estatísticos e da variedade vocabular. Reescrita das produções textuais.

- Retomada dos exercícios estatísticos e da variedade vocabular: os relatórios de estatística lexical são novamente devolvidos aos alunos, que têm em mãos fonte de consulta para reescrita das suas produções textuais, a fim de qualificar seus textos iniciais.
- Reescrita das produções textuais: os alunos são orientados a reescreverem seus textos conforme as anotações individuais da correção, as sugestões de melhoria e, principalmente, os relatórios estatísticos, com resultados observados por eles mesmos, além de opções de variedades lexicais. Acredita-se na maior qualidade dos textos reescritos em função de todos os exercícios realizados e de todas as reflexões feitas. O conjunto das avaliações é entregue: primeira e segunda versões da produção textual e relatório estatístico completo, com todos os exercícios aplicados.

**Parte 6:** Devolução das reescritas das produções textuais dos alunos. Comparação dos textos (primeira e segunda versões). Autoavaliação. Finalização da unidade de ensino.

- Autoavaliação: antes da devolução das avaliações realizadas e entregues pelos alunos, eles são orientados a escreverem sua autoavaliação em relação ao

conjunto de atividades realizadas nessa unidade de ensino. A autoavaliação (ANEXO XIII) deve conter uma nota ou conceito (dependendo da forma de avaliação utilizada pela escola e da pontuação/conceito estipulado pelo professor para a unidade) e uma justificativa para tal, exemplificada com o envolvimento do aluno nas tarefas, suas dificuldades e seus aprendizados. Além disso, haverá também um espaço destinado à avaliação do aluno sobre a unidade de ensino e sobre as atividades desenvolvidas, para que os estudantes possam avaliar a prática, sugerir melhorias e apontar dificuldades presentes na unidade de ensino.

- Devolução das reescritas: após a autoavaliação, as produções e os relatórios estatísticos são devolvidos aos alunos, para que eles tenham conhecimento dos resultados individuais obtidos nessa unidade de ensino.
- Comparação dos textos: com seus textos, primeira e segunda versões, em mãos, acredita-se que os alunos sejam capazes de perceber a diferença e a mudança de uma versão para a outra, depois de estudarem o próprio vocabulário escrito empregado. Então, são instigados a lerem em voz alta, para que os colegas, somente ouvindo, também percebam tais diferenças. Além disso, como atividade para casa, fica a sugestão de que os alunos refaçam a medida estatística de riqueza lexical com as reescritas dos seus textos, para poderem comparar numericamente, em relação a variação vocabular, as diferenças de uma versão para outra.
- Finalização da unidade de ensino: para finalizar, os alunos são instigados a relatarem ao grupo o que acharam da unidade de ensino, o que aprenderam com ela, quais dificuldades encontraram e o que poderia ter sido diferente. Caso tenham vergonha de participar oralmente, é possível ler alguns trechos da autoavaliação da unidade de ensino.

## 7.2 PROPOSTAS DE EXERCÍCIOS (INDEPENDENTES) COM VOCABULÁRIO

A partir de nosso *corpus* de estudo, criamos opções de atividades com vocabulário. Essas atividades, todas independentes, permitem que o aluno explore seu repertório vocabular, apreendendo novas palavras e atribuindo diferentes sentidos para itens já conhecidos.

Assim como considerado na unidade de ensino, essas atividades independentes têm como público-alvo, sobretudo, os alunos do 8º ano do Ensino Fundamental. Um dos motivos é que boa parte de nosso *corpus* de estudo é formado por redações deste ano. Entretanto, assim como na unidade completa, acreditamos que tais exercícios autônomos também podem ser realizados pelos outros anos finais do Ensino Fundamental.

O trabalho com o léxico é fundamental em todas as etapas da vida (não só, mas também escolar) e da formação de um indivíduo, como vimos ao longo deste estudo. Dessa forma, nosso objetivo, com essas atividades, é o de contribuir para que a temática seja abordada em sala de aula de língua materna, possibilitando novos horizontes aos nossos educandos.

Da mesma forma que na unidade de ensino da seção anterior, os exercícios independentes também estão pautados nos princípios teórico-metodológicos da Linguística de Corpus (BERBER SARDINHA, 2004) e da Estatística Lexical (BIDERMAN, 1978; 1998). Entretanto, as atividades são acessíveis aos alunos e professores dos anos finais do Ensino Fundamental, uma vez que eles são o público-alvo de nossas sugestões de atividades.

A seguir estão descritas as atividades de maneira individual, totalizando dez opções. Todas iniciam com breve explicação de cada exercício e, em seguida, o enunciado propriamente dito, seguido das alternativas.

**Público-Alvo:** alunos de 8º ano

**Tempo estimado:** Como as atividades são individuais, é possível realizá-las intercaladas com outros conteúdos e tarefas, não sendo necessárias desenvolvê-las em alguma ordem, nem mesmo todas juntas. Ainda assim, para que se tenha ideia, imaginamos um cenário de um a dois períodos para cada atividade (entre 45min e 100min), individualmente. Sabemos que algumas demandarão um pouco mais de tempo, outras, menos. Mas acreditamos que essa seja a média necessária de tempo para a realização das atividades sugeridas.

Obs.: O total de períodos pode variar conforme o tempo de cada período, que depende da organização da escola e da mantenedora.

**Avaliação:** Contínua, em que o aluno é avaliado principalmente conforme seu interesse, sua participação e seu engajamento nas atividades. Como, nesta parte, idealizamos atividades individuais, não pensamos em entrega de trabalhos, de

relatórios ou de exercícios, mas, sim, no acompanhamento individual do professor no caderno de cada aluno, além do debate das respostas ao final de cada atividade, como forma de compartilhar o conhecimento adquirido, dividindo-o com os colegas e com a turma.

### 7.2.1 Reescritura e contexto

Neste exercício, o aluno terá que reescrever trechos de redações, retirados de nosso *corpus* de estudo, a fim de trocar a palavra sublinhada por outra, que pode ser também uma expressão, de sentido semelhante. Dessa forma, será capaz de perceber quais outras palavras podem desempenhar o mesmo papel, sem, no entanto, produzir grandes alterações no sentido da frase em que está inserida.

EXERCÍCIO: Reescreva o trecho, substituindo a palavra sublinhada por outra palavra ou expressão de seu conhecimento, sem gerar grandes alterações no sentido do trecho/frase: Obs.: Você pode corrigir e/ou reescrever partes que considere necessárias.
a) O Brasil que eu quero é um com menos violência e menos drogas com melhor <u>salário</u> a funcionários públicos e com menos desmatamentos e <u>combater</u> queimadas.
b) Acho que o que eu destacaria na história seria seu crescimento pessoal, seu <u>amadurecimento</u> e sua nova visão da <u>vingança</u> , vendo que não era a melhor solução para seus conflitos pessoais.
c) Tudo começou com uma dupla de ladrões <u>resgatando</u> um presidiário da cadeia.
d) E também a falta de médicos em postos de saúde, hospitais se o governo liberar mais <u>verbas</u> para comprar mais equipamentos [...]
e) Não pelo fato de o garoto <u>despresar</u> a menina e no final eles ficarem juntos. Mas sim pelo motivo de termos algumas <u>recaídas</u> , e no final ter um feliz para sempre.
f) E que animais que são criados em <u>cativeiros</u> sejam liberados e não <u>comercializados</u> [...]
g) Eu iria pagar 1 real por dia pros meus amigos fazerem as <u>filmagens</u> .
h) Bastante professores, sem crimes, não <u>queimar</u> a natureza, não cortar as árvores, poupar a população de fumaças de usinas e de queimadas, sem políticos <u>corruptos</u> .
i) Para resumir um filme de minha vida seria quase uma história de <u>superação</u> tirando a parte que a <u>superação</u> ainda não aconteceu
j) Se você for sair para a rua você tem risco de <u>falecer</u> , mas o Brasil dos meus sonhos não terá esse <u>risco</u> .

### 7.2.2 Relação entre vocábulo e sentido

Neste exercício, o aluno terá que relacionar os vocábulos com seus sentidos possíveis. De um lado, numa coluna, estão os vocábulos, identificados por letras do alfabeto. De outro lado, estão diversos sentidos que esses vocábulos possuem. O aluno precisará relacioná-los.

EXERCÍCIO: Relacione as colunas, ligando as palavras aos seus sentidos possíveis.

Lembre-se de que cada palavra tem mais de um sentido possível:

(a) disputa	( ) remédio, medicamento
(b) destacar	( ) transferência, passagem
(c) perseguição	( ) dever, exercício, tema, trabalho
(d) droga	( ) lento, atrasado, vagaroso
(e) zoar	( ) veículo, viatura, carro
(f) transporte	( ) batalha, briga, combate
(g) lição	( ) justificativa, motivo, razão
(h) retardado	( ) obsessão, assédio
(i) desculpa	( ) sobressair, distinguir-se dos demais
	( ) algo de má qualidade, porcaria
	( ) confusão, tumulto, agitação
	( ) castigo, repreensão, pena, punição
	( ) competição, rivalidade
	( ) indivíduo com problemas de saúde mental
	( ) ressaltar, enfatizar, realçar
	( ) ensinamento, experiência, vivência
	( ) produzir barulho, ressoar
	( ) perdão, absolvição, arrependimento
	( ) entorpecente, alucinógeno, tóxico
	( ) desligar, desunir, separar
	( ) caça, busca, procura
	( ) estúpido, ignorante
	( ) avacalhar, debochar, caçoar

### 7.2.3 Criação de contextos para determinados sentidos

Em complemento ao exercício anterior, em que o aluno relacionou os vocábulos aos seus sentidos possíveis, agora, ele terá que criar frases para o mesmo vocábulo em seus diferentes sentidos.

EXERCÍCIO: Crie frases para cada um dos vocábulos (selecionados do exercício anterior), conforme cada um dos seus sentidos possíveis:
a) disputa:
b) destacar:
c) perseguição:
d) droga:
e) zoar:
f) transporte:
g) lição:
h) retardado:
i) desculpa:
j) conflito:

### 7.2.4 Exploração de sentido de acordo com o contexto

Em uma perspectiva diferente em relação aos exercícios anteriores, nesta atividade, o aluno deve explicar qual sentido da palavra sublinhada conforme seu emprego em um trecho:

Obs.: Os trechos reproduzidos foram adaptados a partir de ocorrências no *corpus* de estudo.

EXERCÍCIO: A partir da leitura dos trechos, explique com as suas palavras qual(is) o(s) sentido(s) da palavra sublinhada:
a) O principal problema do Brasil, nesse exato momento, é a <u>ganância</u> por dinheiro. O homem está matando a natureza por dinheiro!
b) Voltar às aulas é mudar a <u>rotina</u> , acordar cedo, reencontrar os amigos, tudo de novo.
c) O meu filme seria um <u>clichê</u> .
d) Na <u>expectativa</u> , é um grande ano, no qual se pode tudo. Na realidade, é apenas mais um ano.
e) O Brasil que eu quero é um Brasil em que haja mais amor, <u>empatia</u> , respeito, <u>igualdade</u> , felicidade e, tudo que o brasileiro não tem hoje em dia.
f) Fiquei dois anos no time de <u>aspirantes</u> , ganhando muitos títulos, até que eu subi para o time principal e fiquei por lá por mais três anos, ganhando libertadores, etc.
g) Bom, o Brasil que eu quero para o futuro é um Brasil sem <u>imposto</u> , sem crimes e com mais educação.
h) [...] sem ter muitas tarefas que nos <u>sobrecarreguem</u> .
i) Eu gostei muito do <u>otimismo</u> do filho mais novo, ele acreditou o filme inteiro que acharia o tesouro e que não perderia a casa, isso foi bem legal.
j) Mas são apenas palavras, são todas uma <u>ilusão</u> .

### 7.2.5 Conhecido X Desconhecido

Neste exercício, o aluno terá acesso a trechos de redações escolares sobre variados temas. Desses pequenos trechos, deverá marcar, entre as palavras sublinhadas, as que não conhece. Após, deverá procurar no dicionário o significado dos itens desconhecidos, anotando o sentido correspondente ao contexto em que a palavra está.

Obs.: Os trechos reproduzidos foram adaptados a partir de ocorrências no *corpus* de estudo.

EXERCÍCIO: Marque um (X) ao lado de todas as palavras sublinhadas que você não conhece ou não sabe explicar seu sentido. Depois, procure no dicionário o significado dessas palavras. Anote qual sentido de cada uma delas na frase/contexto em que aparecem: (Lembre-se de que o dicionário apresenta diversos sentidos possíveis para cada palavra. Você deve anotar apenas o sentido que se encaixa no contexto):
a) Outra lição é que não podemos <u>julgar</u> ( ) as pessoas pela <u>aparência</u> ( ), por <u>religião</u> ( ), ou por roupa que ela usa.
b) E mais uma vez estou <u>retornando</u> ( ) ao <u>denominado</u> ( ) “ <u>inferno</u> ” ( ), <u>vulgo</u> ( ), escola.
c) O Brasil que eu quero é um Brasil sem <u>guerras</u> ( ) com <u>amor</u> ( ), <u>humildade</u> ( ), <u>felicidade</u> ( ) e <u>alegria</u> ( ).
d) Nisso começa a luta, um <u>adversário</u> ( ) foi até que <u>fácil</u> ( ), ele passou para as <u>oitavas de final</u> ( ) contra Gabriel, que teve suas <u>dificuldades</u> ( ) mas passou, depois Mateus, <u>oponente</u> ( ) forte, luta <u>intensa</u> ( ).
e) A <u>saudade</u> ( ) de ficar sem fazer nada e a <u>indignação</u> ( ) por ter falado <u>incontáveis</u> ( ) vezes: “não tenho <u>nada</u> ( ) pra fazer!” (quando eu <u>realmente</u> ( ) tinha coisas para fazer!).
f) “eu” viveria muitos <u>momentos</u> ( ) <u>inesquecíveis</u> ( ) como uma <u>explosão</u> ( ) do mundo ou uma luta ( ) contra uma <u>enorme</u> ( ) <u>raposa</u> ( ) de 9 caldas ( ).
g) O meio do meu filme traria como <u>conteúdo</u> ( ) a <u>calmaria</u> ( ) após a <u>tempestade</u> ( ). Mostraria as frases <u>motivadoras</u> ( ), porque é aí que mais precisamos de <u>conselhos</u> ( ) e motivações.
h) Os meninos que estavam na busca do <u>tesouro</u> ( ) tiveram que passar por <u>diversas</u> ( ) <u>armadilhas</u> ( ) até chegar em um ponto que os <u>bandidos</u> ( ) chegaram perto deles.
i) Faria como se todos fossem bons em <u>aspectos</u> ( ) <u>diferentes</u> ( ), para mostrar que as <u>diferenças</u> ( ) são boas e que as coisas novas são <u>incríveis</u> ( ), viveria sonhos e os <u>compartilharia</u> . ( )
j) Gabriel tem suas <u>técnicas</u> ( ) <u>escondidas</u> ( ) e ganha com elas, <u>avança</u> ( ) para final contra seu <u>mestre</u> ( ) <u>disfarçado</u> . ( )

### 7.2.6 Variação de formas

Neste exercício, o aluno terá duas colunas. Na primeira, deve atribuir um sentido, uma explicação para palavra, conforme o modelo. Na segunda, deve atribuir a letra correspondente à primeira coluna, relacionando a grafia padrão (1ª coluna) com as grafias encontradas em nosso *corpus* de estudo (2ª coluna).

EXERCÍCIO: As diversas grafias foram retiradas de redações escolares. Relacione todas as grafias de um mesmo item (que podem ter variação de número – singular ou plural – ou de pessoa e tempo, no caso dos verbos), atribuindo à segunda coluna, a letra correspondente da primeira coluna. Em seguida, atribua um sentido, uma explicação para a palavra de grafia culta da primeira coluna, assim como o modelo exemplifica:

<p>(a) água Líquido transparente, sem sabor nem cheiro.</p> <p>(b) assaltos _____</p>	<p>( ) alsatos</p>
<p>(c) muita _____</p>	<p>( a ) agua</p>
<p>(d) estupro _____</p>	<p>( ) munta</p>
<p>(e) família _____</p>	<p>( ) estropo</p>
<p>(f) vivos _____</p>	<p>( a ) aga</p>
<p>(g) haja _____</p>	<p>( ) vissos</p>
<p>(h) bullying _____</p>	<p>( ) bulling</p>
<p>(i) recompensa _____</p>	<p>( ) recompensa</p>
<p>(j) unam _____</p>	<p>( ) bullyng</p>
<p>_____</p>	<p>( ) família</p>
<p>_____</p>	<p>( a ) agúas</p>
<p>_____</p>	<p>( ) estrupos</p>
<p>_____</p>	<p>( ) aja</p>
<p>_____</p>	<p>( ) unão</p>
<p>_____</p>	<p>( ) muntas</p>
<p>_____</p>	<p>( a ) agoas</p>
<p>_____</p>	<p>( ) strupo</p>
<p>_____</p>	<p>( ) mendingo</p>
<p>_____</p>	<p>( ) família</p>
<p>_____</p>	<p>( ) bulyng</p>



### 7.2.7 Palavras de diferentes origens I

Nesta atividade, os alunos receberão frases que contêm palavras de diversas origens estrangeiras que ocorrem no *corpus* de redações, adaptadas para este exercício. Sobre tais palavras, deverão responder e completar algumas informações.

<p>EXERCÍCIO: Sobre cada uma das palavras sublinhadas a seguir, responda:</p> <p>I. Qual seu significado?</p> <p>II. De qual língua herdamos essa palavra?</p> <p>III. Você geralmente usa essa palavra? Em qual contexto?</p>
<p>a) Uma vez, no <u>hotel</u>, essas pessoas tinham se perdido das outras.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>b) Quero um Brasil que todos possam expressar-se livremente, sem se importar com padrões impostos pelas pessoas <u>ignorantes</u>.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>c) No final do filme, meus melhores amigos iriam me visitar no Canadá e assistir ao meu 1º <u>show</u> com a minha banda.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>d) Quero que as pessoas tenham ética e não joguem lixo nos <u>bueiros</u>, nos valões e nas ruas.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>e) Quero que o filme fique o mais real possível, com um diretor qualquer que me escute, o <u>cameraman</u> tem que se profissional, assim como o editor.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>f) Hora de ganhar uma notinha, <u>tchau</u> e até o próximo texto!</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>g) [...] um cara ou guria que sofria <u>bullying</u> ser respeitado e admirado por algo que fez no futuro, por uma descoberta ou até por um <u>recorde</u>.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>

<p>h) E que não sejamos <u>esculachadas</u> por nossas profissões, e sim exaltadas por conseguir efetuar tudo de maneira excepcional.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>i) Eu botaria uma lei que todos os cachorrinhos tenham um <u>chip</u> de identificação com o nome e o endereço de quem o adotou.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>
<p>j) Os pais dele compraram uma <u>picape</u> pra ele.</p> <p>I – significado: _____</p> <p>II – Língua de origem: _____</p> <p>III – Uso: ( ) sim ( ) não</p> <p>Quando? _____</p> <p>Como? _____</p>

## 7.2.8 Palavras de diferentes origens II

Muitas das palavras utilizadas hoje em dia com frequência em nossa língua foram herdadas da língua inglesa. Então, neste exercício, o aluno terá que substituir as palavras sublinhadas por alguma palavra ou expressão equivalente em língua portuguesa.

<p>EXERCÍCIO: Reescreva a frase, substituindo a palavra sublinhada por algum correspondente de sentido semelhante:</p> <p>OBS: Você pode corrigir e/ou reescrever partes que considere necessárias.</p>
a) Não adianta <u>postar</u> nas redes sociais que esta triste pela Amazônia.
b) Ele me ensinaria a jogar <u>videogame</u> .
c) Sábado saio para o <u>shopping</u> ou para casa de meus amigos.
d) Garanto que meu filme pegaria o <u>Oscar</u> .
e) Primeiro aparece o gordo num <u>fliperama</u> tomando iorgute num copo de papel.
f) Eu fui muito bem, marcando dois <u>gols</u> e dando a vitória ao meu time.
g) Quero também mais respeito com os <u>gays</u> , as lésbicas e toda a comunidade LGBT.
h) Depois ia para casa olhar TV, jogar <u>Play</u> ou jogar Free Fire.
i) O gordo, como era chamado pelo os seus amigos, estava comendo numa <u>lanchonete</u> .
j) No final, meu <u>time</u> foi campeão do torneio.

Após a realização do exercício, o professor pode discutir com os alunos a origem dessas palavras, quais sinônimos ou expressões foram indicadas como respostas e até que ponto se consegue ou não utilizar uma forma correspondente em língua portuguesa.

### 7.2.9 Acessibilidade textual

Neste exercício, o aluno fará um *ranking* de complexidade de sinônimos para uma dada situação de uso em uma frase. O aluno deverá atribuir números, que variam de 1 a 4, de acordo com a sua opinião sobre qual das palavras (ou expressões) seria a opção mais facilmente compreendida por ele mesmo ao completar cada contexto.

<p>EXERCÍCIO: Substitua o espaço em branco por cada uma das opções apresentadas e leia a oração novamente. Em seguida, ordene as opções, colocando (1) naquela que torna a leitura da frase mais fácil para você, (2) na próxima mais fácil e, assim por diante. Desta forma, a última opção, com número 4, é a mais difícil para você.</p>
<p>a) Quando ela viu as cartas, não abriu, _____ enviou a todas as pessoas marcadas nas cartas.</p> <p>( ) entretanto  ( ) mas  ( ) todavia  ( ) contudo  ( ) no entanto</p>
<p>b) No carnaval, as pessoas _____ em lugares públicos.</p> <p>( ) urinavam  ( ) mijavam  ( ) faziam xixi  ( ) esvaziavam a bexiga</p>
<p>c) Não se _____ no trabalho e utilizou tecidos de qualidade inferior.</p> <p>( ) empenhou  ( ) dedicou  ( ) esforçou  ( ) aplicou</p>
<p>d) Jennifer _____ e fez a sua última camiseta.</p> <p>( ) acatou  ( ) aceitou  ( ) consentiu  ( ) concordou</p>
<p>e) Minha família tem dias e dias, dias felizes e risonhos e dias tristes sem mesmo olharmos um _____ dos outros.</p> <p>( ) no rosto  ( ) na cara  ( ) na face  ( ) na fuça</p>
<p>f) E arrumar as escolas por que algumas escolas estão precisando de muitos _____.</p> <p>( ) reparos  ( ) consertos  ( ) ajustes  ( ) remendos</p>
<p>g) Graças a minha avó, que me dizia para não desistir dos meus sonho, e sim sempre _____.</p> <p>( ) persistir  ( ) insistir  ( ) prosseguir  ( ) continuar</p>

<p>h) E confio e espero que o Brasil _____.</p> <p><input type="checkbox"/> melhore</p> <p><input type="checkbox"/> evolua</p> <p><input type="checkbox"/> progrida</p> <p><input type="checkbox"/> prospere</p>
<p>i) A minha opinião é que o filme foi bem _____.</p> <p><input type="checkbox"/> elaborado</p> <p><input type="checkbox"/> feito</p> <p><input type="checkbox"/> preparado</p> <p><input type="checkbox"/> produzido</p>
<p>j) Feministas devem ter mais voz de poder para buscar a igualdade e _____ machismo.</p> <p><input type="checkbox"/> abolir o</p> <p><input type="checkbox"/> acabar com o</p> <p><input type="checkbox"/> eliminar o</p> <p><input type="checkbox"/> extinguir o</p>

### 7.2.10 Estatística Lexical

Neste exercício, os alunos receberão um dos textos de nosso *corpus* de estudo (ou outro que o professor queira utilizar, devidamente processado com o *AntConc*), adaptado para esta atividade. Além disso, receberão a lista de palavras desse texto, contendo número total, as formas e as ocorrências de todas as palavras do texto. A partir desses dados, os alunos responderão perguntas sobre o texto.

EXERCÍCIO: Leia o texto. Observe a lista de palavras correspondente a ele: a lista apresenta todas as palavras do texto, em ordem de frequência. Assim, temos que a palavra 'Brasil' ocorre 2 vezes. Você já parou para pensar que algumas palavras a gente usa muito e outras a gente usa pouco? Quais palavras nós mais utilizamos em textos escritos? Depois de pensar nisso, responda as perguntas a seguir:

O Brasil que eu quero é um Brasil com mais educação e menos violência. E com mais respeito com a nossa cor, nacionalidade, porque hoje em dia tem muito preconceito. Nós temos que parar com isso, porque o bullying pode causar a mesma coisa que aconteceu o massacre de Suzano. Com isso, podemos ter uma noção do que o bullying e o preconceito podem causar. Então, temos que pensar melhor no que nós fazemos, para que nada do tipo aconteça de novo. Vamos deixar as pessoas serem do jeito que elas quiserem, para não acontecer isso.

## LISTA DE PALAVRAS

ORDEM	PALAVRA	FREQUÊNCIA
1	que	8
2	com	5
3	o	5
4	do	3
5	e	3
6	isso	3
7	a	2
8	Brasil	2
9	bullying	2
10	causar	2
11	de	2
12	mais	2
13	nós	2
14	para	2
15	porque	2
16	preconceito	2
17	temos	2
18	acontecer	1
19	aconteceu	1
20	aconteça	1
21	as	1
22	coisa	1
23	cor	1
24	deixar	1
25	dia	1
26	educação	1
27	elas	1
28	em	1
29	então	1
30	eu	1
31	fazemos	1
32	hoje	1
33	jeito	1
34	massacre	1
35	melhor	1
36	menos	1
37	mesma	1
38	muito	1
39	nacionalidade	1
40	nada	1
41	no	1
42	nossa	1
43	novo	1
44	noção	1
45	não	1
46	parar	1
47	pensar	1
48	pessoas	1
49	pode	1
50	podem	1
51	podemos	1
52	quero	1
53	quiserem	1
54	respeito	1
55	serem	1
56	Suzano	1
57	tem	1
58	ter	1
59	tipo	1
60	um	1
61	uma	1
62	vamos	1
63	violência	1
64	é	1

PERGUNTAS:
a) Qual é o assunto principal do texto?
b) Quantas palavras há no texto, no total?
c) Deste total, quantas palavras diferentes há no texto?
d) Qual é a palavra mais usada no texto?
e) É verdade que a palavra que a gente mais repete em um texto é a palavra mais importante dele? O que você acha?
f) Quantas palavras ocorrem apenas uma vez no texto?
g) Se você tivesse que dividir as palavras do texto em 4 grupos, como você faria? Leve em consideração os tipos de palavras e as quantidades.

### 7.3 COMENTÁRIOS SOBRE AS PROPOSTAS

Nossas sugestões de atividades com vocabulário escrito têm o objetivo de auxiliar o professor no seu trabalho com o léxico em sala de aula. Em função disso, este capítulo foi escrito para ser compartilhado com os colegas, professores de Língua Portuguesa do Ensino Fundamental.

A primeira parte de nossas sugestões se configura como uma unidade de ensino completa, contendo sugestão de texto, preparação para a leitura, questões de interpretação e de produção baseadas no texto. Também sugerimos um trabalho com o vocabulário escrito dos alunos a partir de suas produções, prevendo uma reescrita conforme o que foi trabalho sobre vocabulário. Nessa parte, pensamos em atividades avaliativas, incluindo uma autoavaliação dos estudos realizados.

A segunda parte sugerida foram atividades individuais, que utilizaram como fonte o nosso *corpus* de estudo. Essas sugestões são possíveis de serem realizadas em variados momentos de estudo, não necessitando utilizar todas elas ou em sequência. Nesta parte, o trabalho é todo focado no vocabulário, seja na relação das palavras com seus diversos sentidos e contextos possíveis, seja no conhecimento que o educando já traz consigo, procurando expandi-lo. Além disso, idealizamos também exercícios com palavras originárias de outros idiomas e com estatística lexical descritiva, todos adaptados ao contexto escolar.

Acreditamos que o professor de língua materna pode ajudar no desenvolvimento da proficiência leitora, interpretativa e compreensiva dos seus alunos por meio de atividades de ensino de vocabulário. Em função disso, sugerimos tais atividades, nesses dois formatos apresentados, a fim de contribuir com os estudos de vocabulário em sala de aula de língua materna. Conforme GIL (2016, p. 462):

O conhecimento de uma palavra não se restringe ao saber um de seus sentidos, o que supostamente permitiria ao aluno prosseguir a leitura de um texto. A competência lexical, como tentamos aqui ilustrar, constitui-se nas relações que o aprendiz estabelece entre a palavra nova e seus conhecimentos prévios. Constitui-se também em um saber mais robusto sobre a palavra: conhecer as frequentes companhias da palavra estudada no nível da língua e do discurso, conhecer os contextos em que a palavra mais frequentemente aparece e saber escolher contextos mais apropriados para seu uso, conhecer propriedades gramaticais da palavra responsáveis pelos sentidos que se pretende estabelecer.

De acordo com a nossa concepção e com base nas referências que adotamos para este trabalho, acreditamos que a proficiência leitora do indivíduo tende a estar ligada ao espectro do seu vocabulário. Assim, quanto mais palavras e características delas um indivíduo souber, melhor será sua compreensão (tanto de textos, como de enunciados, de exercícios, de avisos, de ordens). Segundo Serra (2016, p. 117):

Nesse contexto, é importante ressaltar que, de acordo com Ilari e Cunha Lima (2011, p. 14), em uma sociedade tecnológica com a atual, é exigido dos indivíduos possuir um número razoável de palavras para se comunicar. De acordo com esses autores, "(...) para ter vida normal numa sociedade tecnologicamente avançada como a nossa são necessárias entre 10 e 15 mil palavras (...)", nessa direção, a escola deve ser um espaço em que o aluno possa adquirir, de modo rápido e adequado esse vocabulário que se faz necessário no contexto atual.

Como é na escola, principalmente, que uma criança entra em contato com a leitura e a escrita, pensamos que tal espaço também é privilegiado para o estudo e a expansão do vocabulário que essas pessoas já têm e já trazem de suas experiências de vida anteriores. Ainda conforme Serra (2016, p. 03):

Parte-se do pressuposto que o ensino do vocabulário deve ter como base o texto, a língua em funcionamento, as situações em que o vocabulário possa ser visto em sua função em *in vivo*, situações em que ele é, realmente, um vocabulário (ou vivendo no reino das palavras, conforme dizia Drummond), não uma palavra morta de um dicionário.

Em função disso, criamos atividades em que o vocabulário é (ou será) contextualizado, pois usamos o *corpus* de estudo como fonte, demonstrando o uso do vocabulário explorado, na prática. Tais exercícios foram pensados com o intuito de influenciarem no alargamento do vocabulário dos alunos. Essas sugestões

podem ser aplicadas em sala de aula, com o objetivo de contribuir com a proficiência leitora desses alunos, por meio de exercícios de vocabulário escrito.

Acreditamos que nossa hipótese preliminar de que a expansão de vocabulário escrito pode influenciar positivamente nas capacidades leitoras e interpretativas de um indivíduo poderá ser testada a partir da execução das nossas atividades de ensino de vocabulário, mesmo que combinada com outras. Dessa forma, como perspectiva de nosso trabalho, idealizamos a execução dessas atividades em sala de aula, como forma de percebermos, na prática, o quanto o trabalho com o léxico na escola pode, de fato, auxiliar o desenvolvimento da proficiência desses outros quesitos.



## 8 CONSIDERAÇÕES FINAIS: LIMITES E PERSPECTIVAS

O objetivo principal de nosso estudo foi descrever e analisar o perfil de vocabulário escrito dos estudantes de Ensino Fundamental de escolas públicas de Porto Alegre – RS. Tomamos como referência uma amostra de redações dos nossos alunos, integrantes de uma coleção de textos reunida ao longo de oito anos de docência. A investigação foi realizada por meio de uma análise léxico-estatística descritiva, inspirada nos enfoques basilares de M.T.C. Biderman (1936-2008), com apoio da Lexicologia e da Linguística de Corpus. Além dos elementos quantitativos, naturais a esse tipo de abordagem, buscamos também a contextualização dos resultados em uma perspectiva analítica qualitativa, dedicada ao gênero textual e discursivo identificado como “redação escolar”.

Além de caracterizar padrões de vocabulário escrito, também pretendíamos diferenciar a escrita dos dois grupos de redatores envolvidos: alunos de escola municipal e alunos de escola estadual. Em virtude de nossa experiência docente nessas duas realidades, enxergávamos, subjetivamente, disparidades entre as capacidades de leitura, de interpretação e de compreensão desses estudantes. Em função disso, buscamos identificar em que medida as redações espelhariam, ou não, diferenças por meio do vocabulário escrito.

Entretanto, surpreendentemente, ao final de nossa pesquisa, percebemos que o perfil de vocabulário escrito dos dois públicos tendeu a ser bastante semelhante e igualmente variado. Conforme já discutido ao longo do trabalho, ambos se aproximaram em muitos dos resultados obtidos. A maior distinção ficou por conta de um único quesito: a especificidade de escolhas em itens do tipo *hapax legomenona*. Nesse único aspecto, o vocabulário escrito mostrou-se mais diferenciado entre os dois grupos.

Acreditamos na importância da análise desse grupo de itens, pois, conforme já abordado neste estudo, quanto maior for o repertório lexical de uma pessoa, com mais facilidade ela compreenderá novas palavras, ao deduzir seu significado por meio do contexto. E a extensão desse repertório se apresenta justamente no *hapax*. Então, o estudo desses itens foi de grande valia para esta pesquisa.

Conseguimos caracterizar o perfil de vocabulário desses alunos e obter dados sobre seus padrões de escrita no gênero redação escolar. Ainda assim, ao longo do

percurso, tivemos alguns percalços que nos fizeram repensar materiais e métodos de análise.

A principal dificuldade encontrada neste trabalho foi a de informatizar o *corpus* de estudo, que originalmente é manuscrito e sujeito a incidência de erros ortográficos de vários tipos. Digitar todas as redações escolares, nomeá-las e separá-las de acordo com o ano e a escola foi, com certeza, a parte mais trabalhosa de nosso estudo. Apesar de já termos experiência na confecção e organização de um *corpus* de jornais populares (SILVA, 2011), nos deparamos com novos questionamentos, pois se tratava de textos produzidos por estudantes de Ensino Fundamental. A opção de manter a grafia original das produções, além de requerer desabilitar o corretor automático dos programas de edição de texto, nos fez repensar tal escolha durante o andamento do estudo. Isso porque encontramos variantes escritas da mesma palavra (por exemplo: estrupo/strupo, mendigo/mendingo), o que alterava a contabilização de formas diferentes (*types*).

Assim, refletimos até que ponto, da metade até o final da pesquisa, deveríamos ter uma versão “corrigida” do *corpus* de estudo. Optamos por permanecer com os textos apenas conforme sua grafia original. Isso foi decidido quando nos deparamos com o fato de que tais itens não repercutiam de modo tão significativo sobre a ponderação de palavras-chave, por exemplo, em contraste com outros *corpora*. E, ao final, ainda utilizamos essas variações da escrita dos alunos como um insumo para nossas atividades de ensino de vocabulário, atividades que visam contribuir para reflexão do estudante sobre a escrita de determinadas palavras.

Outra dificuldade encontrada foi a do uso de determinadas ferramentas computacionais, algo que, para profissionais da área de Letras, tende a ser um desafio. Em função disso, além das ferramentas tradicionais de acesso livre da Linguística de Corpus (como o *AntConc*), pudemos contar também com o auxílio dos colegas da Ciência da Computação, do NILC-USP. Seu auxílio foi com a construção de um sistema automático para lematização, baseado na ferramenta UNITEX. Por meio dessa lematização do *corpus* de estudo, afinamos nossos resultados. Foi então que pudemos descobrir que a principal diferença residia no uso dos itens de única ocorrência (*hapax legomena*).

A geração das *keywords* com o *AntConc* mostrou-se um processo bastante complexo. Isso nos levou à utilização indireta do programa *WordSmith Tools*

(SCOTT, 2020). Esse também foi um percalço ao longo de nosso caminho. Para ultrapassarmos essa barreira, contamos com a gentil assistência do pesquisador da UFU, Prof. Dr. Guilherme Fromm. A partir da sua ajuda e da lista de palavras-chave, foi possível, conforme acreditamos, descrever com mais detalhes o nosso *corpus* de estudo.

Outro dos objetivos de nossa pesquisa também era desenvolver práticas de ensino de vocabulário para uso dos professores em sala de aula. Conforme o previsto, idealizamos tais atividades. Muitas delas tiveram como base o *corpus* de estudo. Nossa ideia é compartilhar esse material com os colegas, professores de Língua Portuguesa, para que o ensino do léxico, assim como o de outros conteúdos, também seja priorizado em sala de aula.

Pensamos que tais atividades podem contribuir com a instrumentalização dos professores interessados no tema dos estudos do Léxico, geralmente assoberbados, oportunizando a diversificação do repertório lexical de seus alunos. A partir de um vocabulário escrito mais variado, a tendência é que a compreensão de leitura pelos estudantes, assim como sua escrita se tornem mais abrangentes. Tenderiam, assim, a compreender melhor textos, questões, trabalhos, enunciados, exercícios e quaisquer outras situações que se apresentem em suas vidas. Isso sem contarmos o que significa introduzir esses alunos no tema da escolha de palavras.

Nossa proposta didática, sendo uma sugestão, precisa ser testada, na prática, para que seja possível analisarmos sua influência no desempenho desses alunos. Então, esta é uma das perspectivas de nosso trabalho: a testagem das atividades de ensino de vocabulário propostas.

Outras perspectivas foram sendo desenhadas ao longo dos resultados de nosso trabalho. A partir do que foi possível realizar, podem ser feitos os seguintes estudos:

a) análise do grupo de itens *hapax dislegomena*, que seriam palavras de frequência 2, outro recorte possível em uma lista de frequência de palavras;

b) análise das palavras-chave negativas, que seriam aquelas palavras-chave com frequência significativamente mais alta no *corpus* de referência do que no *corpus* de estudo;

c) análise das palavras-chave exclusivas, que seriam aquelas palavras-chave que ocorrem apenas no *corpus* de estudo, quando em comparação com diversos *corpora* de referência;

- d) análise do tamanho e do padrão das frases;
- e) análise do(s) tipo(s) de verbo(s) mais e menos empregados com seus respectivos complementos.

Tais análises, acreditamos, contribuiriam para ampliar a caracterização do perfil de vocabulário escrito desses alunos. Explorar esses e outros itens de cunho sintático e semântico nos permitiriam agrupar mais dados sobre perfis singulares e comuns de vocabulário.

Nossa caracterização de perfil vocabular de crianças e jovens do Ensino Fundamental, manifestado em suas redações escolares, embora parcial, pode ser aproveitada, por exemplo, para:

1. melhor entender o público/os alunos e suas realidades, a fim de que as aulas, ao se tornarem mais próximas do seu léxico, também se tornem mais prazerosas e produtivas, com aprendizagens significativas;
2. produção de materiais didáticos e de atividades de ensino de línguas;
3. ajuste/adaptação de textos mais complexos e/ou materiais didáticos direcionados a esse público (como, por exemplo, os textos constantes nos experimentos do Museu de Ciências e Tecnologia da PUCRS [MCT-PUCRS], estudadas por colega do grupo de pesquisa, contribuindo para a acessibilidade textual e terminológica);
4. desmistificar o senso comum, sem embasamento científico, de que os alunos de escolas estaduais gaúchas, de maneira geral, apresentam melhor desempenho e maior repertório vocabular do que os alunos de escolas municipais;
5. evidenciar que o CorPop, *corpus* construído também por integrante do nosso grupo de pesquisa, de fato, serve como uma referência válida de português popular escrito culto para diferentes finalidades de investigação.

Por fim, acreditamos que a pesquisa científica pode e deve contribuir para que se possa refletir sobre a atividade do profissional docente, acolhendo suas experiências na academia. Foi por meio das experiências de sala de aula que surgiram nossas inquietações sobre a origem de, em tese, diferentes competências dos estudantes. E foi também por meio de atividades de ensino com a redação escolar, incluindo seus materiais de apoio e modos de realização, que conseguimos equacionar nossos questionamentos. Em função disso, entendemos que a pesquisa

e a docência precisam caminhar sempre juntas, estabelecendo parcerias que produzam frutos para ambas.

## REFERÊNCIAS BIBLIOGRÁFICAS

AMORIM, M. B.; LACERDA, M. P. M. Semântica e Gramática. **Cadernos de Letras da UFF - GLC**, n. 27, p. 181-185, 2003.

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

ANTUNES, I. **Aula de Português: encontro e interação**. São Paulo: Parábola, 2003.

ANTUNES, I. **Território das palavras: estudo do léxico em sala de aula**. São Paulo: Parábola, 2014.

ASSUNÇÃO, E. D. **Ampliação vocabular: glossário de textos do livro didático de língua portuguesa "Vontade de saber português" do 9º ano**. 2015. 118 f. Dissertação (Mestrado em Letras) – Universidade Federal de Uberlândia, Uberlândia, 2015.

BAKER, M. A corpus-based view of similarity and difference in translation. In: ARDUINI, S.; HODGSON, R. (org.). **Translating Similarity and Difference**. Manchester: St. Jerome, 2004, p. 1-18.

BAKHTIN, M. Os gêneros do discurso. In: BAKHTIN, M. **Estética da criação verbal**. São Paulo: Martins Fontes, 2003. p. 261-306.

BARBEIRO, L. F. Diversidade Lingüística na Escola Portuguesa. In: **Projecto Diversidade Lingüística na Escola Portuguesa (ILTEC)**. Lisboa: ILCTEC [s.n., entre 2003 e 2007]. Disponível em: <[http://www.iltec.pt/divling/\\_pdfs/cd2\\_vocabulario\\_quadro-geral.pdf](http://www.iltec.pt/divling/_pdfs/cd2_vocabulario_quadro-geral.pdf)>. Acesso em: 05 abr. 2018.

BARBOSA, M. A. Contribuição ao estudo dos aspectos da tipologia de obras lexicográficas. **Ciência da Informação**, Brasília, v. 24, n. 3, [sp], 1995. Disponível em: <<http://revista.ibict.br/ciinf/article/view/572/573>> Acesso em: 18 mar. 2021.

BARREIROS, P. N.; BARREIROS, L. L. S. O vocabulário da ditadura militar nos panfletos de Eulálio Motta. **Filol. Linguíst. Port.**, São Paulo, v. 17, n. 2, p. 385-420, jul./dez. 2015. Disponível em: <<http://dx.doi.org/10.11606/issn.2176-9419.v17i2p385-420>>. Acesso em: 21 ago. 2020.

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

BERBER SARDINHA, T. Linguística de Corpus: histórico e problemática. **D.E.L.T.A.**, São Paulo, v. 16, n. 02, p. 323-367, 2000.

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

BERBER SARDINHA, T.; BARBARA, L. Frequência e uso de estrangeirismos ingleses no português brasileiro: Um estudo baseado em corpus. **Rev. bras. linguist. apl.**, Belo Horizonte, v. 5, n. 1, p. 97-114, 2005. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1984-63982005000100006&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1984-63982005000100006&lng=en&nrm=iso)>. Acesso em: 09 mar. 2020.

BERBER SARDINHA, T. O banco de palavras-chave como instrumento de identificação de palavras-chave exclusivas no programa WordSmith Tools. **the ESPecialist**, vol. 27, n. 1, p. 1-19, 2006. Disponível em: <<https://revistas.pucsp.br/index.php/esp/article/view/1626/1045>>. Acesso em: 09 mar. 2020.

BERBER SARDINHA, T. **Pesquisa em linguística de Corpus com WordSmith Tools**. Campinas: Mercado das Letras, 2009.

BERBER SARDINHA, T.; MOREIRA FILHO, J. L.; ALAMBERT, E. O Corpus Brasileiro. **Comunicação apresentada em 13º Congresso Brasileiro de Língua Portuguesa e 4º Congresso Internacional de Lusofonia**, PUCSP, São Paulo, SP, 30 de abril de 2010.

BERTONHA, F. H. C.; ZAVAGLIA, C. Dinamismo lexical nas redes sociais: contribuições para Lexicografia. **Calidoscópico**, Unisinos, v. 05, n. 03, p. 407-416, setembro-dezembro, 2017.

BIBER, D. Representatividade em planejamento de corpus. Trad. Paula Marcolin. **Cadernos de Tradução**, Porto Alegre, n. 30, p. 11-25, jan-jun, 2012.

BIDERMAN, M. T. C. **Estatística linguística**. Alfa, São Paulo, v. 11, p. 117-128, 1967.

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

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

BIDERMAN, M. T. Dimensões da Palavra. **Filologia e Linguística Portuguesa**, n. 02, p. 81-118, 1998b.

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

BRASIL. Lei no 9.394, de 20 de dezembro de 1996. Estabelece as diretrizes e bases da educação nacional. **Diário Oficial da União**, Brasília, DF, 23 dez. 1996. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/leis/L9394.htm](http://www.planalto.gov.br/ccivil_03/leis/L9394.htm)>. Acesso em: 14 mai. 2020.

BRASIL. Lei nº 11.274, 6 de fevereiro de 2006. Altera a redação dos arts. 29, 30, 32 e 87 da Lei nº 9.394, de 20 de dezembro de 1996, que estabelece as diretrizes e bases para a educação nacional, dispondo sobre a duração de 9 (nove) anos para o ensino fundamental, com matrícula obrigatória a partir dos 6 (seis) anos de idade. **Diário Oficial da União**, Brasília, DF, 7 fev. 2006a. Disponível em: <[http://www.planalto.gov.br/ccivil\\_03/\\_ato2004-2006/2006/lei/11274.htm](http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/11274.htm)>. Acesso em: 14 mai. 2020.

BRASIL. Ministério da Educação. Secretaria de Educação Básica. **Com direito à palavra**: dicionários em sala de aula / [elaboração Egon Rangel]. – Brasília: Ministério da Educação, Secretaria de Educação Básica, 2012.

BRASIL. Secretaria de Educação Fundamental. **Parâmetros curriculares nacionais**: terceiro e quarto ciclos do ensino fundamental: língua portuguesa/ Secretaria de Educação Fundamental. Brasília : MEC/SEF, 1998. 106 p.

CÂMARA, M. J., J. **Princípios de Lingüística Geral**. 4. ed. Rio de Janeiro: Livraria Acadêmica, 1967.

CARECHO, J.; SOARES, R.; FERNANDES, A. Tecnologias Digitais numa Experiência de Aprendizagem Autônoma de Léxico da Língua Alemã. **Diacrítica**, vol. 34, nº. 01, p. 248-274, 2020. Disponível em: <<https://doi.org/10.21814/diacritica.355>>. Acesso em: 24 abr. 2020.

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

COLOMBO, R. C.; CÁRNIO, M. S. Compreensão de leituras e vocabulário receptivo em escolares típicos de Ensino Fundamental I. **Codas**, São Paulo, vol. 30, nº 4, 2018. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S2317-17822018000400306&lng=en&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2317-17822018000400306&lng=en&nrm=iso)>. Acesso em: 25 abr. 2020.

CORREIA, M. Léxico, Corpora e Tradução: Diálogos com a Sociedade. *In*: Seminário de Estudos Linguísticos da UNESP (12: 2020: Araraquara, SP). **Caderno de resumos [do] XII Seminário de Estudos Linguísticos da UNESP**: 16 a 18 de setembro de 2020, Araraquara -SP / [Organização de Gabriel Guimarães Alexandre, Maria Angélica Deângeli e Talita Serpa]. São José do Rio Preto: UNESP/IBILCE, 2020. 195 p.

COSERIU, E. **Gramática, semántica, universales**. Editorial Gredos: Madrid, 1987. 268 p.

CÚRCIO, V. R. Estudos estatísticos de textos literários. **Texto Digital**, Florianópolis, ano 2, n. 2, 2006. Disponível em: <<https://periodicos.ufsc.br/index.php/textodigital/article/view/1382/1080>>. Acesso em 19 ago. 2020.



DAMIM, C. P. **Parâmetros para uma avaliação do dicionário escolar**. 2005. 230 f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2005.

DAMIM, C.; PERUZZO, M. S. Uma descrição dos dicionários escolares no Brasil. **Cadernos de Tradução**, Florianópolis, v. 2, n. 18, p. 93-113, abr. 2006. ISSN 2175-7968. Disponível em: <<https://periodicos.ufsc.br/index.php/traducao/article/view/6981/6450>>. Acesso em: 19 ago. 2020.

DARGEL, A. P. T. P. **O ensino do vocabulário nas aulas de Língua Portuguesa: da realidade a um modelo didático**. 2011. 290 f. Tese (Doutorado em Linguística e Língua Portuguesa) – Faculdade de Ciências e Letras, Universidade Estadual Paulista Júlio de Mesquita, Araraquara, 2011.

DIAS, E. Reflexões sobre o ensino do léxico na escola fundamental. In: MARTINS, E. S.; CANO, V. M.; MORAES FILHO, W. B. (org.). **Léxico e morfofonologia: perspectivas e análises**. Uberlândia: EDUFU, 2006.

DUNCAN, J. A. **Frequency Dictionary of Portuguese Words**. Stanford: Stanford University, 1972.

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

ESTOPÀ, R. (org.). **Hablamos de lengua com niños y niñas**. Barcelona: Editora GRAÓ, 2006.

EVERS, A. **Processamento de língua natural e níveis de proficiência do português: um estudo de produções textuais do exame Celpe-Bras**. 2013. 174 f. Dissertação (Mestrado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2013.

EVERS, A. **A redação engaiolada: padrões lexicais e ensino de redação em cursos pré-vestibulares populares**. 2018. 229 f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

FERREIRA, H, M.; VIEIRA, M, S, P. O trabalho com o léxico em sala de aula: desafios para o ensino de língua materna. **Revista Letras Raras**, v. 2, n. 2, p. 19-33, 2013.

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.

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: 15 abr. 2020.

FINATTO, M. J. B. Vocabulário em redações de vestibulandos: contribuições da léxico-estatística textual, palavras-signo e palavras-em-uso. *In*: REBELO, L. S.; FLORES, V. N. (org.). **O Texto de Vestibular em Perspectiva**. Comissão Permanente de Seleção (COPERSE/UFRGS). Porto Alegre: Editora UFRGS, 2014. p. 171-183.

FINATTO, M. J. B. Trabalho com pequenas e grandes amostras textuais: levantamento de terminologias na área Pneumopatias Ocupacionais. *In*: ISQUERDO, A. N.; DAL CORNO, G. O. M. **A Ciências do Léxico, Lexicologia, Lexicografia, Terminologia**. Campo Grande: Editora da UFMS, 2018.

FORTES, C. A. **As estratégias de revisão de textos nos anos iniciais do Ensino Fundamental**. 2016. 48 f. Trabalho de Conclusão de Curso (Graduação em Pedagogia) – Faculdade de Educação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

FROMM, G. A construção e análise de corpora para alimentação de um banco de dados terminográfico: um exemplo. **Domínios de Lingu@gem**, v. 2, n. 1, 31 jan. 2011. Disponível em: <<http://www.ileel.ufu.br/guifromm/wp-content/uploads/2014/05/11483-42686-1-PB.pdf>>. Acesso em: 12 ago. 2020.

GIAMMATTEO, M. ALBANO, H. El “valor” de la palabra: el léxico como elemento integrador de las competencias lingüística y comunicativa. *In*: GIAMMATTEO, M. ALBANO, H. (org.). **El léxico**. De la vida cotidiana a la comunicación cibernética. Buenos Aires: Biblos, 2012. p. 19-49.

GIL, B. D. Ensino de vocabulário e competência lexical. **Gragoatá**, Niterói, vol. 21, n. 40, 2016, p. 445-464.

GRAMA, D. F. **Uma análise lexicográfica dos elementos coesivos sequenciais do Português para a elaboração de uma proposta de definição**: um estudo com base em corpus. 2016. 370 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Uberlândia, Uberlândia, 2016.

GUEDES, P. C.; SOUZA, J. M. D. Leitura e escrita são tarefas da escola e não só do professor de português. *In*: CONCEIÇÃO, I. C. B.; SOUZA, J. V.; SHAFFER, N. O.; GUEDES, P. C.; KLUSSENER, R. (org.). **Ler e escrever**: compromisso de todas as áreas. 9 ed. Porto Alegre: Editora da UFRGS, 2011.

HARTMANN, N. S. **Adaptação lexical automática em textos informativos para o Ensino Fundamental**. 2020. 180 p. Tese (Doutorado em Ciências –

Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2020.

ILARI, R. Uma nota sobre redação escolar. **ALFA**, Assis, v. 22-23, p. 82-101, out., 1978. Disponível em: <<https://periodicos.fclar.unesp.br/alfa/article/view/3587/3356>>. Acesso em: 09 mar. 2021.

INAF/ INSTITUTO PAULO MONTENEGRO. **Indicador de Alfabetismo Funcional (Inaf) - Principais Resultados**. [S.l.]. 2018. Disponível em: <[https://acaoeducativa.org.br/wp-content/uploads/2018/08/Inaf2018\\_Relat%C3%B3rio-Resultados-Preliminares\\_v08Ago2018.pdf](https://acaoeducativa.org.br/wp-content/uploads/2018/08/Inaf2018_Relat%C3%B3rio-Resultados-Preliminares_v08Ago2018.pdf)> Acesso em: 09 mar. 2021.

KADER, C. C.; RICHTER, M. G. Linguística de corpus: possibilidades e avanços. **Instrumento: R. Est. Pesq. Educ.**, Juiz de Fora, v. 15, n. 1, jan./jun. 2013. p. 14-23.

LOPES, L.; FERNANDES, P.; VIEIRA, R. Estimating term domain relevance through term frequency, disjoint corpora frequency -tf-dcf. **Knowledge-based Systems**, p. 237-249, 2016.

LORENTE, M. A Lexicologia como ponto de encontro entre a Gramática e a Semântica. Tradução de Schreiber e Padilha. *In*: ISQUERDO, A.N.; KRIEGER, M.G. (org.). **Ciências do Léxico**, v. II, 2004, p. 19-30.

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.

MACIEL, A. M. B. Linguística de Corpus: bases teórico-metodológicas. Colóquio Nacional de Letras em diálogo e em contexto: rumos e desafios, 2002, Porto Alegre. **Anais...** Porto Alegre: Programa de Pós Graduação em Letras, 2003. 1CD-Rom.

MAGALHÃES, C. M.; NOVODVORSKI, A. A chavidade na análise de estilo em tradução: um estudo baseado em corpora paralelos espanhol/português. X Encontro de Linguística de Corpus, 2012, Belo Horizonte. **Anais do X Encontro de Linguística de Corpus**: aspectos metodológicos dos estudos de corpora. Belo Horizonte: FALE/UFMG, 2012. v. 1. p. 294-313.

MAINARDES, J. Ciclos de formação. *In*: OLIVEIRA, D. A.; DUARTE, A. M. C.; VIEIRA, L. M. F. (org.). **DICIONÁRIO**: trabalho, profissão e condição docente. Belo Horizonte: UFMG/Faculdade de Educação, 2010. CDROM.

MARTINS, T. B. F.; GHIRALDELO, C. M.; NUNES, M, G.; OLIVEIRA JÚNIORS, O. N. Readability formulas applied to textbooks in Brazilian-Portuguese. **Relatório Técnico**, São Carlos: ICMC/USP, 1996.

- NASCIMENTO, R. I. **O vocabulário dos estudantes universitários**: um estudo com base em redações de vestibular. 2001. 159 f. Dissertação (Mestrado em Letras) – Universidade Federal de Mato Grosso do Sul, Três Lagoas, 2001.
- NASCIMENTO, R. I.; ISQUERDO, A. N. Análise estatística e sociolingüística do vocabulário de estudantes: desvendando desigualdades sociais e problemas no ensino. **50º Seminário do Grupo de Estudos Lingüísticos do Estado de São Paulo**, 2002, São Paulo. Resumos. São Paulo: Contexto, 2002. p. 242.
- NASCIMENTO, R. I.; ISQUERDO, A. N. Frequência de palavras: um diagnóstico do vocabulário de redações de vestibular. **Alfa**, São Paulo, v. 47, p. 71-84, 2003.
- NAUJORKS, J. C. **Leitura e Enunciação**: princípios para uma análise do sentido na linguagem. 2011. 153 f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2011.
- NOVODVORSKI, A.; BOCORNY FINATTO, M. J. Linguística de Corpus no Brasil: uma aventura mais do que adequada. **Letras & Letras**, Uberlândia, v. 30, n. 2, p. 7-16, 2014.
- NOVODVORSKI, A. As palavras-chave empo/tempo: um estudo empírico-descriptivo em corpus literário traduzido. **Letras & Letras**, Uberlândia, v. 32, p. 170-193, 2016.
- OLIVEIRA, T. A.; SILVA, E. G. O.; SILVA, C. O.; ARAÚJO, L. A. M. **Tecendo Linguagens – Língua Portuguesa 8º ano**. 4. ed. São Paulo: IBEP, 2015.
- PAIS, A. P.; SARDINHA, M. G. G. A. Os caminhos do Ensino e da Aprendizagem do Vocabulário: Intercultura, Corpus Textual e Lexicultura. **Congresso Internacional de Literatura Infantil e Mundo Globalizado**, 2011. Disponível em: <<https://core.ac.uk/download/pdf/62718289.pdf>>. Acesso em: 19 ago. 2020.
- PALAVRA**. In: DICIONÁRIO Michaelis. Disponível em: <<https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/palavra/>>. Acesso em: 19 ago. 2020.
- PASQUALINI, B. F. **CorPop**: um corpus de referência do português popular escrito do Brasil. 2018. 250 f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.
- PIRES, D. R.; TUMOLO, C. S. INSTRUÇÃO DE VOCABULÁRIO EM L2. **Diacrítica**, Minho, vol. 34, n. 01, p. 225-247. Disponível em: <<https://doi.org/10.21814/diacritica.303>>. Acesso em: 24 abr. 2020.
- POLGUÈRE, A. **Lexicologia e semântica lexical**: noções fundamentais. Tradução de Sabrina Pereira de Abreu. São Paulo: Contexto, 2018.

RECKSKI, L. J. Concordâncias, listas de palavras e palavras-chave: o que elas podem nos dizer sobre a linguagem?. **Lit. lingüíst.**, Santiago, n. 16, p. 249-261, 2005. Disponível em: <[https://scielo.conicyt.cl/scielo.php?script=sci\\_arttext&pid=S0716-58112005000100014&lng=es&nrm=iso](https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0716-58112005000100014&lng=es&nrm=iso)>. Acesso em: 17 abr. 2020.

RIO GRANDE DO SUL, Secretaria de Educação do Estado. **Referenciais Curriculares para Educação Básica: área de Linguagens e Códigos**. Vol 1. Porto Alegre: SERS/DP, 2009.

RIO-TORTO, G. O Léxico: semântica e gramática das unidades lexicais. *In*: ATHAYDE, M. F. (ed.). **Estudos sobre léxico e gramática**. Coimbra : CIEG/FLUC, 2006. p. 11-34.

RODRIGUES, M. I. M. **Vocabulário e (des)motivação para leitura**. Dissertação (Mestrado em Português Língua Não Materna) – Universidade Aberta, Lisboa, 2014. Disponível em: <[https://repositorioaberto.uab.pt/bitstream/10400.2/3496/1/TMPLNM\\_Marialsab elRodrigues.pdf](https://repositorioaberto.uab.pt/bitstream/10400.2/3496/1/TMPLNM_Marialsab%20elRodrigues.pdf)>. Acesso em: 10 nov. 2020.

SANCHEZ, A.; CANTOS, P. **CUMBRE – Curso de Español**. Madri, SGEL, 1996.

SANTOS, D. Corporizando algumas questões. *In*: TAGNIN, S.; VALE, O. A. (org.). **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 41-66.

SCOTT, M. **WordSmith Tools versão 8**, Stroud: Lexical Analysis Software, 2020.

SERRA, L. H. O ensino de vocabulário na sala de aula: reflexões e práticas para a produção de textos na educação básica. **Revista Afluente: revista eletrônica de Letras e Linguística**, v. 01, p. 01-19, 2016.

SOUZA, J. V. de P.; ROCHA, C. F. Análise dos vocábulos de maior chavidade em “A mensagem”, conto clariceano traduzido para as línguas inglesa e espanhola. **Mosaico**, São José do Rio Preto, v. 18, n. 1, p. 452-479, 2019.

SIGNOR, R. C. F. **Os gêneros do discurso (resenha)**. Revista Gatilho (PPGL/UFJF. Online), v. 7, p. s/p, 2008. Disponível em: <<https://periodicos.ufjf.br/index.php/gatilho/article/view/26929>>. Acesso em: 01 dez. 2020.

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

SILVA, B. R. Vocabulário escrito de estudantes e de materiais didáticos do Ensino Fundamental nas redes públicas municipal e estadual de educação. **Revista GTLex**, p. 1-26, 1 abr. 2020.

SILVA, E. B. Perfil lexical de redações de futuros professores de língua inglesa: um estudo de caso. **LINGUAGEM & ENSINO**, Pelotas, v. 19, p. 113-136, 2016.

TAGNIN, S. E. O. **O jeito que a gente diz**: combinações consagradas em inglês e português. Barueri: DISAL, 2013.

TCACENCO, L. M.; SILVA, B. R.; FINATTO, M. J. B. Acessibilidade textual e terminológica. **Revista GTLex**, v. 3, n. 2, p. 197-224, 19 fev. 2020.

TONELI, P. M. AS PALAVRAS FUNCIONAIS DO PORTUGUÊS BRASILEIRO: UMA ANÁLISE SOB A ÓTICA DA INTERFACE SINTAXE-FONOLOGIA. *In*: CELSUL, 2010, PALHOÇA. **IX Encontro do Círculo de Estudos Linguísticos do Sul - CELSUL**. Palhoça: Editora da Unisul, 2010.

VIANA, V. Linguística de corpus: conceitos, técnicas e análises. *In*: VIANA, V.; TAGNIN, S. E. O. (org.). **Corpora no ensino de línguas estrangeiras**. 1. ed. São Paulo: HUB Editorial, 2010. v. 1. p. 25-95.

**ANEXO I – UMA REDAÇÃO COMPLETA DE CADA ESCOLA****Redação sobre a temática 1 “A vida como filme” – Escola Municipal**A história da minha vida

Seria um filme bem grande, com momentos tristes e felizes, e os atores e atrizes seria<sup>m</sup> meus amigos e minha família. Eu contaria tudo, desde de quando eu nasci, até os dias de hoje.

Eu destacaria várias coisas, como eu falei pela primeira vez, quando dei o meu primeiro passo, a primeira vez que fui a escola, destacaria também as minhas formaturas (ABC, minha conclusão de término do ensino fundamental I e a minha formatura do curso de inglês).

Com certeza, eu não deixaria passar nada, a principal atriz seria minha vó, minha mãe e minha prima, pois, foi elas que mediram e fizeram tudo o que eu sou hoje. Meu pai seria o ator principal, eu eliminaria muitas coisas tristes, pois passei muitos.

Talvez um dia eu vá pra Disney e isso com certeza eu não vou deixar passar, ~~mas~~, se eu não for, contarei sobre minhas outras viagens, contarei também o dia em que eu sai de Belém e vim morar no Rio Grande do Sul (Porto Alegre).

Contarei também sobre o momento mais triste da minha vida, pois não pode faltar um pouco de tristeza em um filme, foi o dia em que minha vó partiu.

Vou falar também como foi a minha experiência na minha primeira viagem de avião. Estas são algumas das coisas que aconteceriam no meu filme



**Redação sobre a temática 1 "A vida como filme" – Escola Estadual**

Se eu pudesse fazer um filme da minha vida ia ser filme de comédia, tragédia e drama... Pois minha vida é agitada, iria ser muito engraçado porque eu iria contar todos os meus "podres" da minha vida. Ia ser o filme que contava minha adolescência, sobre as festinhas que eu fui, as minhas histórias românticas. Falar dos meus crush's e das minhas amigas.

Quero que o filme fique mais real o possível, com um diretor qualquer que me escute, o camera-man tem que ser profissional, assim que nem o editor. No máximo 6 meses de filmagem, depois de terminar o filme quero fazer um jantar e uns dias antes de lançarem o filme, quero fazer uma pré-estreia, com tudo que tenho direito.



## ANEXO II – TEMAS E PROPOSTAS DAS PRODUÇÕES

TEMAS: são cinco temas diferentes.				
TEMA 1: A vida como filme	TEMA 2: Um sonho possível	TEMA 3: Voltar de férias é...	TEMA 4: Os Goonies	TEMA 5: O Brasil que eu quero...

PROPOSTAS
<p><b>PROPOSTA 1 – A vida como filme</b> Agora é a sua vez: escreva um texto contando como seria o filme da sua vida. Explique quais seriam os atores e atrizes do seu filme, o que aconteceria na sua história, o que você destacaria e o que você eliminaria em seu roteiro. Seja criativo e não copie as ideias da escritora! Boa produção!</p>
<p><b>PROPOSTA 2 – Um sonho possível</b> Escreva um texto contando qual a lição que se pode tirar do filme ‘Um sonho possível’ e o que podemos aprender com ele. Não faça resumos! Resumos não serão corrigidos! Exponha sua(s) opinião(s), justificando-a(s), em, no mínimo, 20 linhas. Seja criativo e boa produção!</p>
<p><b>PROPOSTA 3 – Voltar de férias é...</b> Agora é a sua vez: produza um texto semelhante ao da estudante Nathália. Diga o que é, para você, ‘voltar de férias’. Seja criativo e não copie as ideias da escritora! Boa produção! Mínimo: 20 linhas.</p>
<p><b>PROPOSTA 4 – Os Goonies</b> Escreva um texto contando a história do filme ‘Os Goonies’ (de maneira geral), expondo sua opinião sobre o filme e justificando suas ideias. Bom trabalho!</p>
<p><b>PROPOSTA 5 - O Brasil que eu quero</b> Agora é a sua vez: elabore um texto com, no mínimo, 15 linhas, respondendo a pergunta: que Brasil que eu quero para o futuro? Cite exemplos, busque valores importantes para você e justifique sua(s) opinião(s). Seja criativo! Boa produção!</p>

## ANEXO III – CORPORA DE CONTRASTE: TEXTOS DE APOIO

# A vida como filme

**S**e pudéssemos fazer de nossa vida um filme seria, não tenhamos dúvida, um grande filme. Uma superprodução, dessas que custam US\$ 100 milhões e batem recordes de bilheteria. Ou, ao contrário, um filme alternativo, como aqueles do Cinema Novo ("Uma câmera na mão, uma idéia na cabeça"): não faz diferença. O filme de nossa vida seria, independentemente de orçamento, uma obra-prima.

Para começar, criaríamos uma boa história. Uma história com começo, meio e fim, uma história capaz de dar à existência um sentido – coisa que ela nem sempre tem – e um final feliz, coisa que ela raramente tem. Na nossa história, eliminaríamos todos os momentos chatos, aborrecidos. Em nosso filme, nunca passaríamos longas horas numa fila. Em nosso filme, o trânsito nunca congestionaria. Em nosso filme, todas as pessoas seriam amáveis e cordiais. E bonitas. Ah, isso o nosso filme teria: belas atrizes, inteligentes atores. Uma dessas atrizes ou um desses atores faria, com inspiração, o nosso papel – sob nossa orientação. Seríamos os diretores de nosso filme. Com amplos e ilimitados poderes. Mais ou menos os poderes que tem Deus. Talvez não fizéssemos milagres, mas teríamos acesso a tudo aquilo que a tecnologia pode propor-

cionar, em termos de imagem, de som. Poderíamos, se quiséssemos, usar efeitos especiais. Por exemplo: o chefe que nos incomoda apareceria no filme apenas para se transformar num sapo (e um sapo que nem mesmo beijo de princesa recuperaria para a forma original).

O nosso filme concorreria ao Oscar, claro. Cartas marcadas: no momento em que o apresentador estivesse dizendo "E o Oscar vai para..."; nós, na plateia, já estaríamos levantando, sorridentes, já estaríamos olhando ao redor com ar de triunfo, já estaríamos nos encaminhando para o palco, para ali receber sob uma chuva de aplausos, o troféu a que sempre fizemos jus. E depois nos daríamos até ao luxo de fazer um discurso convencional – aqueles idiotas discursos do Oscar – agradecendo aos nossos pais, ao nosso cônjuge, aos nossos filhos, aos nossos amigos, ao homem da banca de jornais. E sairíamos dali consagrados.

Isso tudo se a vida fosse um filme. Mas a vida, infelizmente, não é um filme. Ela dura muito mais do que um filme e tem muito menos sentido do que um filme. Só saberemos se receberá ou não um prêmio no Dia do Juízo Final, que, até o momento, ainda não foi marcado, o que representa um grande transtorno em termos de planejamento da produção. A vida não é um filme. E é por isso que preferimos os filmes.

Manoel Scliar 24/2003

# CONCURSO CULTURAL

## Voltar de férias é... retomar a rotina



Zero Hora publica hoje o terceiro dos 10 textos selecionados na terceira edição do concurso cultural Voltar de Férias É..., que convidava estudantes de todas as idades a escrever livremente sobre o tema. Nathália Prestes Bosak,

14 anos anos, é a autora do terceiro texto da série, que teve a participação de mais de cem estudantes de todo o Estado. Leia o que diz Nathália, que está no 1º ano do Ensino Médio do Colégio João XXIII, em Porto Alegre:

### NATHÁLIA PRESTES BOSAK

Voltar de férias é retomar o contato (ao vivo e a cores) com a galera do colégio. É poder conversar, abraçar e olhar nos olhos de todos com quem, nas férias, eu só falava via internet. É retomar os papos da aula de história e as intermináveis discussões na aula de geografia. E também é descobrir as fofocas do verão inteiro em um único dia!

É retomar a rotina perdida depois do último dia de aula: ser arrancada da cama antes das sete horas, tomar café da manhã, aprontar-se para a escola, ir para aula, lanchar correndo no recreio, voltar para a aula e depois que o sinal tocar, correr para não perder a carona para casa; almoçar, ir para a aula de dança, de inglês, de música, fazer as tarefas, tomar banho, jantar, entrar no computador, ir dormir e sonhar com aquele dia, nas férias, em que eu reclamei que não tinha nada para fazer.

Voltar de férias é entrar na sala de aula, no primeiro dia, com aquela cara de envergonhada que eu, sem sucesso, até tento disfarçar – já que quem está ali me cumprimentando

são só meus queridos colegas (os mesmos desde o jardim de infância). É, depois do primeiro momento, retomar a confiança e o meu lugar na terceira classe da fileira perto da janela, com vista privilegiada para todos os cantos da sala. É retomar os estudos com um baita desafio pela frente: como será esse tal de Ensino Médio? E perceber que dizer “tchau” para o Ensino Fundamental não é só uma conquista, mas o fim da minha vida social em todas as semanas em que eu tenho prova.

É perceber que, a partir de agora, o negócio é sério. É me acostumar com o fato de que “ves-



tibular” será a palavra que eu mais vou escutar nos três próximos anos. É aproveitar muito o único dia letivo sem provas, trabalhos ou professores de mau-humor. É não se importar com a fila da cantina. É conversar em aula sem ser advertida. É curtir esse primeiro dia como se fosse o último, mesmo sabendo que ainda tenho um ano inteiro pela frente.







**NÍLSON SOUZA**  
nilsouza31@gmail.com

## Futuro

- Alô, Rede Globo! Alô, Brasil!  
- Sim, já coloquei o celular na horizontal, deitado, como diz o Bonner para não deixar dúvidas, pois - triste realidade! - há muitos brasileiros que não sabem distinguir horizontal de vertical (de pé). Meu nome é Nilson (não posso esquecer esta parte), e estou falando de Porto Alegre, a minha cidade natal. Não, a imagem que aparece atrás de mim não é a Usina do Gasômetro, nem a Ponte do Guaíba, a Catedral Metropolitana ou qualquer outro desses lindos monumentos da nossa Capital. É uma escola pública. Chama-se Dr. Ferreira de Abreu e fica no bairro Sarandi, na Zona Norte, antigo reduto de operários, hoje um aglomerado urbano gigantesco formado por muitas vilas regulares e vários loteamentos clandestinos. Foi nessa escola que aprendi a ler, a escrever, a fazer contas e a diferenciar horizontal de vertical. Sem pagar um centavo, até mesmo porque meu pai era leiteiro e mal ganhava o suficiente para garantir o leitinho da gurizada, com perdão do trocadilho. Mas todos nós, eu e meus irmãos, estudamos nessa escola, que não era de tijolo à vista como é hoje, nem jamais esteve pichada como está. Na época, a garotada aprendia em casa que não se toca em bem alheio e que o que é público é de todos. Nossa escolinha era uma construção de madeira antiga, em outro ponto do bairro, tão precária, que bebíamos água tirada de um poço no centro do pátio. Água encanada, naquele tempo, nem pensar. Mas o Ferreirinha, como carinhosamente o chamávamos, foi a nossa janela para o mundo. Com um grupo de professores abnegados - que certamente também ganhavam mal, mas acho que recebiam em dia, pois jamais houve interrupção de aulas por greve -, aprendemos tudo o que precisávamos para seguir adiante na vida. Muitos de nós - e agora falo de todos os alunos - saímos capacitados para as etapas seguintes da educação formal e, posteriormente, para a universidade. Mesmo os que não continuaram estudando receberam da escola base intelectual e moral suficiente para construir seus próprios futuros. Isto me faz lembrar que o motivo desta gravação imaginária é dizer que Brasil eu quero para o futuro. Que tal um Brasil voltado para o futuro do futuro, em que as crianças recebam proteção (e limites) em casa, ensino qualificado na escola e, na vida, o desafio permanente de construir um país melhor?

- Como? Passei dos 15 segundos? Meu vídeo não vai entrar? Sem problema. Alguém, com maior capacidade de síntese, certamente contará uma história parecida para lembrar que só com educação de qualidade haverá futuro para o nosso país.

**GAUCHAZH**  
1500 70000  
www.gazetaonline.com.br

### Futuro

- Alô, Rede Globo! Alô, Brasil!  
- Sim, já coloquei o celular na horizontal, deitado, como diz o Bonner para não deixar dúvidas, pois - triste realidade! - há muitos brasileiros que não sabem distinguir horizontal de vertical (de pé). Meu nome é Nilson (não posso esquecer esta parte), e estou falando de Porto Alegre, a minha cidade natal. Não, a imagem que aparece atrás de mim não é a Usina do Gasômetro, nem a Ponte do Guaíba, a Catedral Metropolitana ou qualquer outro desses lindos monumentos da nossa Capital. É uma escola pública. Chama-se Dr. Ferreira de Abreu e fica no bairro Sarandi, na Zona Norte, antigo reduto de operários, hoje um aglomerado urbano gigantesco formado por muitas vilas regulares e vários loteamentos clandestinos. Foi nessa escola que aprendi a ler, a escrever, a fazer contas e a diferenciar horizontal de vertical. Sem pagar um centavo, até mesmo porque meu pai era leiteiro e mal ganhava o suficiente para garantir o leitinho da gurizada, com perdão do trocadilho. Mas todos nós, eu e meus irmãos, estudamos nessa escola, que não era de tijolo à vista como é hoje, nem jamais esteve pichada como está. Na época, a garotada aprendia em casa que não se toca em bem alheio e o que é público é de todos. Nossa escolinha era uma construção de madeira antiga, em outro ponto do bairro, tão precária que bebíamos água tirada de um poço no centro do pátio. Água encanada, naquele tempo, nem pensar. Mas o Ferreirinha, como carinhosamente o chamávamos, foi a nossa janela para o mundo. Com um grupo de professores abnegados - que certamente também ganhavam mal, mas acho que recebiam em dia, pois jamais houve interrupção de aulas por greve -, aprendemos tudo o que precisávamos para seguir adiante na vida. Muitos de nós - e agora falo de todos os alunos - saímos capacitados para as etapas seguintes da educação formal e, posteriormente, para universidade.

Mesmo os que não continuaram estudando receberam da escola base intelectual e moral suficiente para construir seus próprios futuros. Isso me faz lembrar que o motivo desta gravação imaginária é dizer que Brasil eu quero para o futuro. Que tal um Brasil voltado para o futuro do futuro, em que as crianças recebam proteção (e limites) em casa, ensino qualificado na escola e, na vida, o desafio permanente de construir um país melhor?

- Como? Passei dos 15 segundos? Meu vídeo não vai entrar? Sem problema. Alguém, com maior capacidade de síntese, certamente contará uma história parecida para lembrar que só com educação de qualidade haverá futuro para nosso país.

(Nilson Souza - Zero Hora 07/02/18)

## ANEXO IV – CORPORA DE CONTRASTE: TEXTOS DOS LIVROS DIDÁTICOS

### A cara vida moderna

Meu primeiro celular parecia um tijolo. Difícil de carregar. Pior ainda, de funcionar. A linha vivia com sinal de ocupado. Mesmo assim era um luxo! Lembro quando liguei pela primeira vez para minha amiga Vera:

- Estou em Brasília, no meu celular — contei.
- Também quero um! — ela gritou, entusiasmada.

De novidade, tornou-se essencial. Agora esses aparelhos são mínimos, fotografam, tocam músicas e acessam a internet. Viver sem um é estar desconectado. No fim do mês vem a conta. Sempre me assusto! As operadoras oferecem pacotes. E de pacote em pacote às vezes eu me sinto enbrulhado! Compro por puro entusiasmo uma série de serviços que não uso depois! Um amigo meu tem três celulares. Durante um jantar, falava em todos ao mesmo tempo, enquanto eu tentava conversar. Imagino a conta!

A cada dia inventam algo que imediatamente se torna indispensável. Impossível encontrar um adolescente que não sinta necessidade de um laptop. Se não tem, voa para uma lan house. A internet ficou tão importante quanto as calças que estou vestindo. O laptop de um jovem ator quebrou às vésperas de ele sair em turnê pelo país com um espetáculo. Está desesperado.

— Vou perder meu contato com o mundo!

É verdade! E-mails, redes de relacionamento e blogs são vitais para boa parte das pessoas. Tudo isso custa: o orçamento cresce em eletricidade, conexões de banda larga e equipamentos — os avanços são rápidos, é preciso renovar sempre. Falando em avanços: um amigo formou uma excelente coleção de clássicos de cinema em vídeo. Jogou fora e iniciou outra ao surgir o DVD. Agora veio o Blu-ray. O coitado quase explodiu de tão estressado! Mas é impossível permanecer com o equipamento antigo. Em pouco tempo some das lojas. Toca comprar tudo de novo!

A TV por assinatura tornou-se um sonho de consumo. E os televisores em si? Todo dia fico sabendo de uma tela maior, mais fina e com melhor imagem. Sem falar nos eletrodomésticos, mais e mais sofisticados. Quando comprei o meu primeiro freezer, há muito tempo, um amigo riu:

— Para que uma coisa dessas?



Hoje ninguém dispensa um freezer. Qualquer item da vida pode se sofisticar: faz-se café expresso em casa, sorvete, iogurte e até pão. Ninguém tem tudo, é fato. Mas todo mundo tenta ter algum novo e fantástico produto!

Passada a garantia, é difícil consertar qualquer aparelho. O preço raramente compensa. E logo quebra de novo, mesmo porque muitos técnicos de antigamente perderam o pé nos digitais!

Viver ficou muito mais caro. Antes eu parava o carro na rua, agora é Zona Azul ou estacionamento particular; os cinemas aumentaram o valor dos ingressos porque investem em tecnologia; cabeleireiros sofisticaram os produtos: banho em cachorro é melhor no pet shop; é essencial um cartão de crédito, mas vem a anuidade. Além de um bom plano de saúde, é ideal também um de aposentadoria. Tenho certeza: daqui a pouco descobrirei algo absolutamente essencial de cuja existência até agora não tinha o menor conhecimento!

Mas os salários não subiram na mesma proporção. No passado era mais fácil cortar gastos. Agora, não. Muitas despesas não podem mais sair do orçamento. Contatos profissionais, bancários e muitos serviços públicos acontecem através de celulares e da internet. Já conheci gente com falta de dinheiro para comer, mas sem poder abdicar do celular!





[...] Foram cinco tardes de muito estudo. E Reinaldo tinha um objetivo a atingir. Por isso, tratou de prestar atenção às lições de Ricardo e raciocinar. Resultado: aprendeu direitinho o que queria.

– Hoje, vou dar um show! – garantiu a uma colega, no início da aula, alguns dias depois.

Era só esperar a professora chamá-lo, pensou. Vinte minutos passados, bateu a impaciência. Será que a professora tinha esquecido dele? No quadro-negro, outro menino tinha acabado de grifar os sujeitos das orações que Márcia pedira.

– Parabéns! – a professora cumprimentou os acertos.

Reinaldo levantou a mão. Olhava para a primeira oração escrita no quadro. Ela era:

*O time da escola venceu os visitantes por dois a zero.*

O sujeito da oração, *O time da escola*, tinha sido sublinhado.

– Só pra confirmar, professora... – Reinaldo esclareceu o motivo do aparte. – *Time* é o núcleo do sujeito?

– Exatamente – concordou Márcia.

– Eu sabia! – exclamou o garoto, para marcar que sabia mesmo.

Algumas caras de espanto, outras de gozação se viraram para ele. Com o canto do olho, Reinaldo pescou o olhar que lhe interessava. Parece que Valéria tinha gostado da exibição.

Eduardo não gostou nem um pouco. Resolveu se intrometer, falando com a professora.

– Esse negócio de núcleo do sujeito a gente ainda não aprendeu – disse, numa queixa.

Gol contra. Reinaldo aproveitou e continuou o show. Falou de peito cheio para os colegas:

– O núcleo é a palavra central do sujeito. A mais importante de todas que fazem parte do sujeito.

No caso, trata-se de *time*. *Time* é o elemento principal. O vencedor dos visitantes.

E se voltou para Valéria, lembrando:

– Com a modesta participação dos meus passes para o Chico.

A classe estava de queixo caído. Será que Reinaldo tinha tomado chá de enciclopédia? Até Márcia estava calada. Como todas as atenções continuassem nele, Reinaldo soltou mais um exemplo:

– Naquela outra oração, *A professora de Matemática não veio hoje*, *professora* é o núcleo do sujeito. É a palavra que exerce o papel central.

– Como é que dá para garantir isso? – perguntou a Regininha, lá no fundo da classe.

Reinaldo não vacilou:

– Se a gente tirar a palavra *professora*, a oração fica até sem sentido.

– A... *de Matemática não veio hoje* – repetiu a Regininha em voz alta.

– Fica sem sentido mesmo! – concordou Valéria.

Amigo de Eduardo, Filipe sussurrou alguma coisa em seu ouvido. O garoto levantou a cabeça. Seus olhos brilharam.

– E você sabe dizer, Reinaldo, se esse sujeito é simples ou composto? – perguntou Eduardo, certo de que colocava o colega contra a parede.

– Quero ver ele sair dessa – comentou Filipe, apertando a mão do amigo.

Muita gente ficou de orelha em pé para escutar a resposta. Márcia ainda não ensinara a classificação do sujeito. Para a maioria da turma, esse assunto não podia ser coisa fácil.

– É sujeito simples – Reinaldo respondeu, superior. – Só tem um núcleo, *professora*. Aliás, como eu já disse.



Todo mundo se voltou para Márcia, esperando a confirmação.

– Muito bem, Reinaldo! – a professora estava mesmo surpresa. – Continue assim.

A essa altura, o garoto queria mesmo esbanjar.

– Aí no quadro, só tem uma oração com sujeito composto. É: *Meu irmão e a prima de Maria foram ao cinema* – Reinaldo foi em frente. – Dá licença, Márcia?

Chegou até o quadro, grifando as palavras *irmão* e *prima*.

– Estas são palavras principais do sujeito, são seus núcleos. Quando o sujeito de uma oração tem mais de um núcleo, ele é um sujeito composto. Certo, professora?

- Certíssimo!

O sinal tocou. A confusão da saída começou. Reinaldo largou o giz. Foi buscar o material na sua carteira. Antes parou ao lado de Valéria. Respirou fundo.

- Não era má ideia um cineminha hoje à tarde... - convidou.

- Se a minha mãe deixar - a menina sorriu. - Me telefona...

Na volta para casa, Teleco estranhou o silêncio de Eduardo.

- O que é que está acontecendo com você? - quis saber curioso.

- Não dá pra explicar - resmungou Eduardo, carrancudo. - Na sua idade, você não vai entender.

A diferença de idade dos dois era pequena. Mas Eduardo a usava, quando queria evitar que Teleco se intrometesse nas suas coisas. O irmão menor ficava bravo:

- Deixa de ser crica...

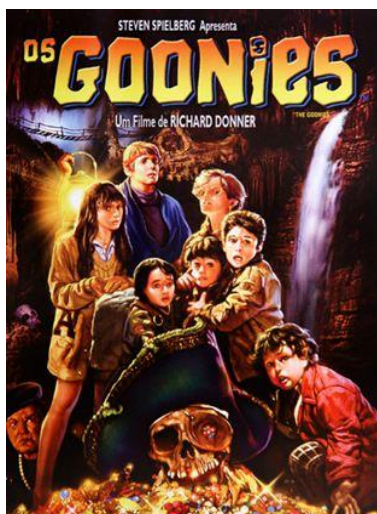
Eduardo precisava desabafar:

- É a Valéria, você sabe... - falou vagamente.

- Se soubesse não tava perguntando... - retrucou Teleco, impaciente.

OLIVIERI, Antônio Carlos. *Um bom sujeito*. Belo Horizonte: Formato, 1997.

## ANEXO V – FILMES USADOS COMO “APOIO” (RESUMO, IMAGEM)



### OS GOONIES – 1985

Com os prédios de seu bairro estando prestes a ser demolidos, o que forçará a mudança de todos os residentes do local, um grupo de garotos resolve organizar uma cerimônia de despedida do local. Quando descobrem um legítimo mapa do tesouro, capaz de torná-los ricos e evitar a destruição de suas casas, Os Goonies resolvem partir em uma grande aventura.

Disponível em: <http://www.adorocinema.com/filmes/filme-1051/>

Acesso em: 03/01/2021



### UM SONHO POSSÍVEL – 2009

Michael Oher (Quinton Aaron) era um jovem negro, filho de uma mãe viciada e não tinha onde morar. Com boa vocação para os esportes, um dia ele foi avistado pela família de Leigh Anne Tuohy (Sandra Bullock), andando em direção ao estádio da escola para poder dormir longe da chuva. Ao ser convidado para passar uma noite na casa dos milionários, Michael não tinha ideia que aquele dia iria mudar para sempre a sua vida, tornando-se mais tarde um astro do futebol americano.

Disponível em: <http://www.adorocinema.com/filmes/filme-132048/>

Acesso em: 03/01/2021



## ANEXO VI – MATERIAL DE APOIO À UNIDADE DE ENSINO: FIGURAS E TEXTOS



**Frases/Textos:**

“O título da Copa do Brasil será decidido em Porto Alegre, no Estádio Beira Rio.”

(número total de palavras: 15 / número de palavras diferentes: 15 / palavras repetidas: 0 - ou seja, todas as palavras têm a mesma frequência: 1)

“Hoje visitei a velha estação de trem. A estação estava coberta de abandono. Porque, hoje, as pessoas parecem não dar bola para construções velhas como essa da estação de trem. Muitas pessoas só querem saber de coisas futuras. Dão bola só para coisas novas.”

(número total de palavras: 44 / número de palavras diferentes: 31 / palavras repetidas: de (4), estação(3), a(2) , hoje(2), trem(2), pessoas (2), coisas (2), bola(2), para(2), só(2) – ou seja, as palavras com mais ocorrência no *corpus* são de e estação, em seguida, todas de frequência 2 listadas)

“Os empresários têm encontrado certos problemas para contratar mão-de-obra especializada, nesses últimos meses. O problema da mão-de-obra é consequência de um problema maior: os altos níveis de desemprego constatados algum tempo atrás. Enfrentando problemas para conseguir empregos nas fábricas a que estavam acostumados, dedicaram-se a outras atividades, criando, para as indústrias, o problema de não encontrar pessoas acostumadas a funções específicas. Demorará ainda algum tempo para que este problema seja solucionado”.

(número total de palavras: 71 / número de palavras diferentes: 54 / palavras repetidas: para(4), problema(4), de(3), a(3), algum(2), o(2), mão-de-obra(2), os(2), problemas(2), que(2), tempo(2) – ou seja, as palavras com mais ocorrência no *corpus* são para e problema, em seguida, de e a, depois, todas de frequência 2 listadas)

### **Pesquisas:**

- PorPopular: Projeto Padrões do Português Popular Escrito, com duas fases, desenvolvido por um grupo de pesquisa da UFRGS, cujo objetivo era a caracterização do léxico e da linguagem como um todo em um texto que é feito, em tese, de modo mais simplificado, para ser compreendido com facilidade por pessoas de um determinado grupo social e econômico. O *corpus* principal de estudo foram: o jornal Diário Gaúcho (fase1) e o jornal Massa (fase2).

Site: <http://www.ufrgs.br/textecc/porlexbras/porpopular/>

- Sem medo do *internetês*: pesquisa de pós-graduação sobre a escrita de jovens brasileiros tal como acontecia no site de relacionamentos *Orkut*. Esse tipo de escrita, apesar das alterações gráficas que a identificam, é bastante integrado aos padrões lexicais do português brasileiro. A pesquisa buscou mostrar e provar que, nesse meio, a escrita não rompe com a escrita “normal” e formal, como muitos acreditam que é a escrita da internet, mas é sim, capaz de se adaptar e se transformar, já que a língua é viva e está sempre em constante modificação.

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

### **Corpora:**

- CorPop: *corpus* do português popular brasileiro escrito, compilado a partir de textos selecionados com base no nível de letramento médio dos leitores do país.

Os textos selecionados para o CorPop são:

(1) textos do jornalismo popular do Projeto PorPopular (jornal *Diário Gaúcho*), consumido maciçamente pelas classes C e D, em que está o leitor médio brasileiro;

(2) textos e autores mais lidos pelos respondentes das últimas edições da pesquisa *Retratos da Leitura no Brasil*;

(3) coleção "É Só o Começo" (adaptação de clássicos da literatura brasileira para leitores com baixo letramento, adaptação esta realizada por linguistas);

(4) textos do jornal *Boca de Rua*, produzido por pessoas em situação de rua, com baixa escolaridade e baixo letramento; e

(5) textos do *Diário da Causa Operária*, imprensa operária brasileira produzida também por pessoas dentro da faixa média de letramento do país.

Site: <http://www.ufrgs.br/textecc/porlexbras/corpop/>

- NURC: O acervo do Projeto NURC-RJ (Projeto da Norma Urbana Oral Culta do Rio de Janeiro), disponível on-line, trata-se de *corpus* de linguagem oral, pois é constituído de entrevistas gravadas nas décadas de 70 e 90 do século XX, num total de 350 horas, com informantes com nível superior completo, nascidos no Rio de Janeiro e filhos de pais preferencialmente cariocas. É usado, principalmente, como referência para estudos da variante culta da língua portuguesa, também como contraste em pesquisas dos mais variados assuntos.

Site: <http://www.nurcrj.lettras.ufrj.br/>

## **ANEXO VII – MATERIAL DE APOIO À UNIDADE DE ENSINO: QUESTÕES DE INTERPRETAÇÃO**

### **Questões de Interpretação**

- a)** Qual é o assunto principal do texto?
  
- b)** Qual é a linguagem utilizada no texto? (gírias? palavras desconhecidas? linguagem coloquial? linguagem regional?) Dê exemplos retirados do texto.
  
- c)** Conforme o texto, o que eliminaríamos em nossa história?
  
- d)** Segundo o autor, em nosso filme, o que nunca aconteceria?
  
- e)** De acordo com a crônica, o que aconteceria conosco, se nosso filme concorresse ao Oscar?
  
- f)** Conforme Scliar, como seriam as pessoas em nosso filme?
  
- g)** O texto termina com: “A vida não é um filme. E é por isso que preferimos os filmes.” O que o autor quis dizer neste trecho?
  
- h)** Como seria o filme de sua vida? Por quê?
  
- i)** O que você eliminaria do seu roteiro?
  
- j)** Ao que você daria destaque?

## ANEXO VIII – MATERIAL DE APOIO À UNIDADE DE ENSINO: PROPOSTA DE PRODUÇÃO TEXTUAL

### Proposta de Produção

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)

## ANEXO IX – MATERIAL DE APOIO À UNIDADE DE ENSINO: MODELO DE RELATÓRIO

Nome:

Turma:

### Relatório estatístico

Título do texto:

Número total de palavras do texto:

Número de palavras diferentes repetidas:

Lista de palavras do texto (ordem alfabética)


Lista de palavras do texto (ordem de frequência)


Palavras lexicais mais frequentes no texto, contexto e possível substituição:

Palavra lexical	Contexto	Possível substituição

Riqueza Lexical:

Gráfico de frequência de palavras:

**ANEXO X – MATERIAL DE APOIO À UNIDADE DE ENSINO: RIQUEZA  
LEXICAL**

$$\text{Riqueza Lexical} = \frac{\text{types} \times 100}{\text{tokens}}$$

Legenda:

*Tokens*: número total de palavras de um texto (frase, parágrafo ou o que estiver sendo analisado).

Conforme Bisognin (2009), “*tokens*: contagem de palavras ocorridas; cada palavra conta como uma ocorrência, mesmo que seja repetida.” (BISOGNIN, 2009, p. 46).

*Types*: número de palavras diferentes repetidas de um texto (frase ou o que estiver sendo analisado).

Conforme Bisognin (2009), “*types*: número de palavras diferentes.”(BISOGNIN, 2007, p. 46).

Exemplo: A professora da Camila disse que a Camila era a melhor aluna da turma.

*Tokens*: 14 (a, professora, da, Camila, disse, que, a, Camila, era, a, melhor, aluna, da, turma)

*Types*: 10 (a, professora, da, Camila, disse, que, era, melhor, aluna, turma)



## **ANEXO XI – MATERIAL DE APOIO À UNIDADE DE ENSINO: TEXTOS COM DIFERENTES RIQUEZAS**

### **TEXTO 1: A vida como filme (Moacyr Scliar) – crônica (ANEXO III)**

### **TEXTO 2: Resumo de artigo científico - resumo**

Trata-se da produção de cogumelos com resíduo proveniente da cerveja. Com intuito de contribuir para sustentabilidade do planeta foi realizado um estudo que objetivou-se em converter o bagaço de malte, resíduo produzido pelas cervejarias, em nutrientes para produção de cogumelos. Os cogumelos são seres pertencentes ao reino Fungi com grande potencial para humanidade, pois podem ser utilizados em diversas aplicações. Criou-se um meio de cultivo para produção de cogumelos com a adição do bagaço de malte, que foi comparado com um meio convencional e com bagaço de malte puro, avaliando o crescimento micelial dos fungos: *Pleurotus ostreatus* e *Pleurotus sajor-caju* Singer. Com base nesse estudo observou-se um equilíbrio no meio de cultura com de cultura com adição de bagaço de malte, pois o mesmo apresentou tanto um bom crescimento quanto um vigor intermediário entre os meios comparados. Comparou-se também a eficiência biológica da produção do cogumelo *Pleurotus Sajor-caju* Singer com o meio de cultivo com adição de bagaço de malte e com o meio convencional utilizando o cultivo axênico. Diante dos resultados de eficiência biológica novamente o meio de cultura com adição de bagaço de malte surpreendeu ao apresentar um valor de eficiência biológica um pouco maior que o do meio convencional. Através dessas comparações foi possível concluir que o bagaço de malte pode ser utilizado para produção de cogumelos e que é muito promissor, podendo fornecer mais resultados interessantes dependendo da espécie a ser cultivada, ambiente ou variações em sua composição.

**Referência: SILVA, VÍTOLA, SILVA, 2019.**

TEXTO 3: Lista de compras – lista



Site: <https://www.elo7.com.br/lista-de-compras-arquivo-digital/dp/D86765>

TEXTO 4: Redação escolar – redação

O Brasil que eu quero é um Brasil sem corrupção, sem políticos roubando, sem brigas e sem violência.

O Brasil que eu quero é um Brasil sem perigo quando se anda de noite na rua.

O Brasil que eu quero é um Brasil sem crianças passando fome, sem desemprego e sem pobreza.

O Brasil que eu quero é um Brasil sem drogados assaltando pra comprar drogas.

O Brasil que eu quero é um Brasil sem falta de professores nas escolas, sem vandalismo nas escolas e sem pinchação nas escolas.

O Brasil que eu quero é um Brasil sem falta de vaga nas escolas perto da nossa casa.

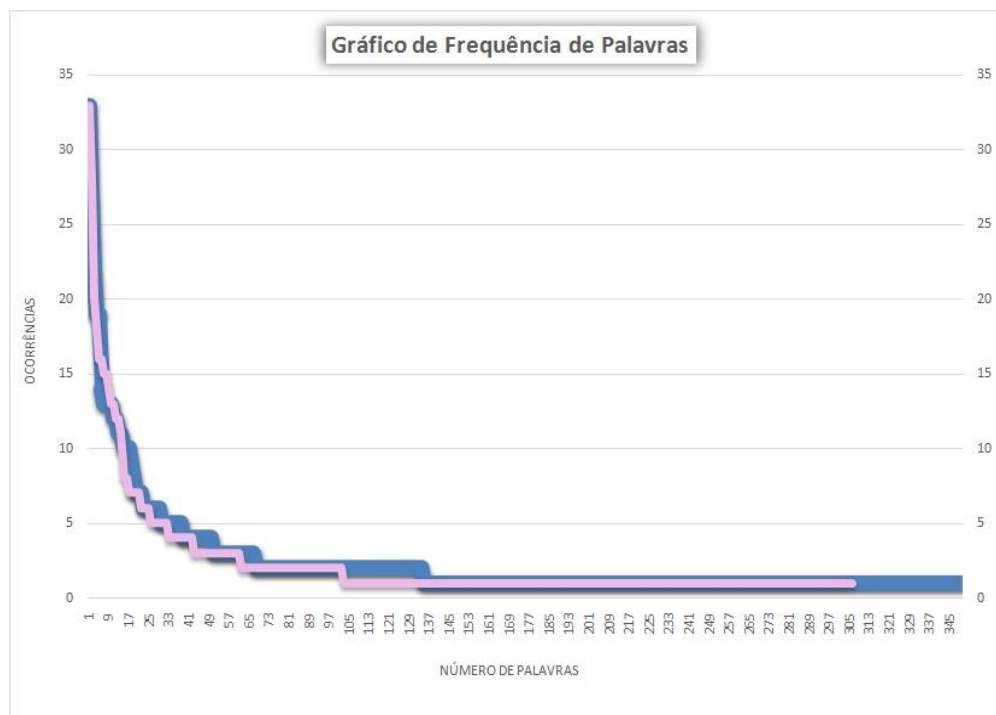
O Brasil que eu quero é um Brasil sem poluição, sem alagamento e sem doenças.

Esse é o Brasil que eu quero.

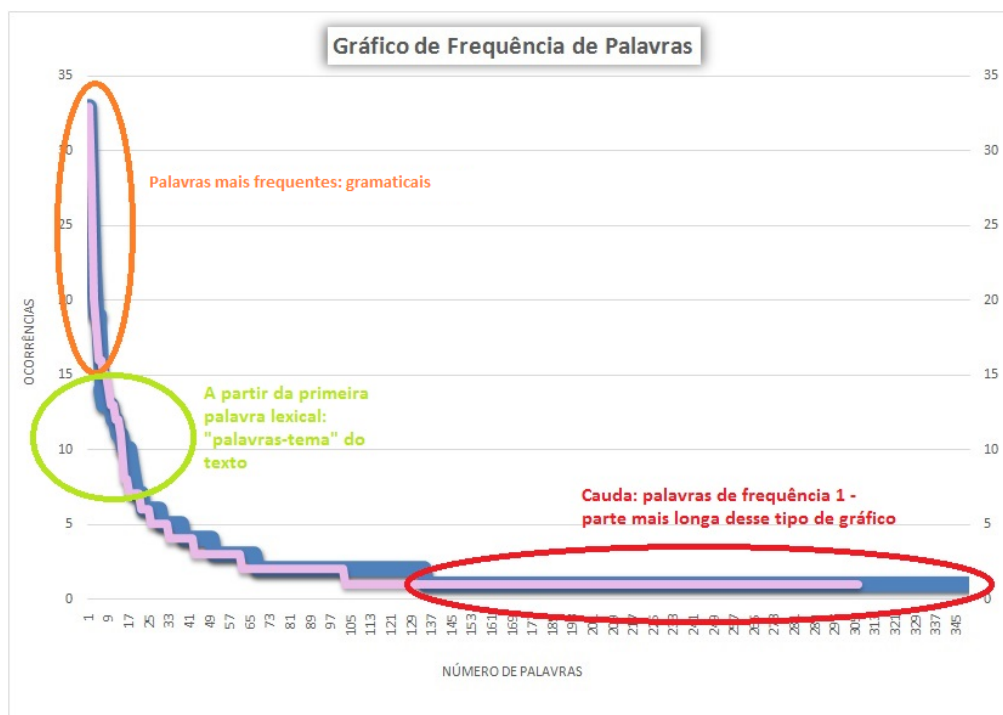
Diferentes riquezas lexicais

	TEXTO 1	TEXTO 2	TEXTO 3	TEXTO 4
<b>Riqueza Lexical</b>	52,3%	50,2%	79,4%	36,4%

## ANEXO XII – MATERIAL DE APOIO À UNIDADE DE ENSINO: MODELO DE GRÁFICO



### Demonstrativo das camadas



**ANEXO XIII – MATERIAL DE APOIO À UNIDADE DE ENSINO:  
AUTOAVALIAÇÃO**

**Modelo de autoavaliação**

Nome:

Turma:

**Autoavaliação**

**Parte 1: pessoal**

Nota/conceito: \_\_\_\_\_

Justificativa (neste espaço, justifique por que você atribuiu essa nota/conceito para sua atuação nas aulas durante essa unidade de ensino, exemplifique com seu envolvimento nas tarefas, suas dificuldades e seus aprendizados):

---

---

---

---

---

---

---

**Parte 2: unidade de ensino**

O que você achou da unidade de ensino como um todo? Deixe sua opinião, sempre justificando-a:

---

---

---

---

---

Qual parte dessa unidade de ensino você mais gostou? Por quê?

---

---

---

Qual parte dessa unidade de ensino você não gostou? Por quê?

---

---

Você tem sugestão do que poderia ser diferente nesse projeto? Suas ideias são muito bem-vindas! Escreva-as aqui:

---

---

---

---

---

---

---

---

## ANEXO XIV – WORDLIST DO CORPUS DE ESTUDO

#Word Types: 4113

#Word Tokens: 26424

#Search Hits: 0

1	1211	que	36	99	dia
2	1124	e	37	97	ter
3	716	o	38	96	ou
4	620	de	39	95	ser
5	527	a	40	94	todos
6	483	eu	41	93	quando
7	458	não	42	91	vida
8	426	para	43	86	depois
9	426	um	44	86	isso
10	359	os	45	82	bom
11	349	mais	46	79	porque
12	338	as	47	78	tudo
13	293	com	48	76	coisas
14	267	é	49	75	futuro
15	233	no	50	73	seria
16	227	brasil	51	72	me
17	217	se	52	71	como
18	209	uma	53	69	então
19	204	do	54	69	escola
20	204	em	55	69	também
21	201	por	56	67	amigos
22	198	pessoas	57	67	pra
23	176	na	58	63	das
24	158	eles	59	63	foi
25	157	minha	60	62	dos
26	156	quero	61	60	até
27	143	da	62	58	professores
28	141	filme	63	57	nas
29	139	tem	64	56	queria
30	132	muito	65	55	melhor
31	124	ele	66	55	meus
32	122	mas	67	54	bem
33	120	sem	68	54	fazer
34	106	meu	69	54	só
35	102	casa	70	52	escolas

71	51	seu	111	31	mundo
72	51	sua	112	30	acho
73	50	são	113	30	aulas
74	50	você	114	30	ruas
75	49	era	115	29	estudar
76	49	tenha	116	29	etc
77	48	mãe	117	29	pelo
78	47	menos	118	28	anos
79	46	ela	119	28	ficar
80	45	família	120	28	pode
81	45	nos	121	28	pois
82	45	nós	122	28	temos
83	43	hoje	123	27	ajudar
84	43	tinha	124	27	aula
85	43	vai	125	27	crianças
86	41	ir	126	27	iria
87	40	coisa	127	27	mudar
88	40	tesouro	128	27	outros
89	40	voltar	129	27	sempre
90	39	estão	130	27	tambem
91	39	muitas	131	26	ao
92	39	onde	132	26	elas
93	38	dinheiro	133	26	estava
94	38	lixo	134	26	muita
95	38	pai	135	26	pessoa
96	38	pais	136	26	pouco
97	37	nada	137	26	tempo
98	36	cada	138	26	vezes
99	35	ano	139	25	dele
100	35	assim	140	25	gordo
101	35	nem	141	25	irmão
102	34	seus	142	25	mim
103	33	final	143	25	tenham
104	33	sair	144	24	dias
105	33	todo	145	24	já
106	32	educação	146	24	outras
107	32	esse	147	24	restaurante
108	32	lá	148	24	sobre
109	32	mesmo	149	23	bandidos
110	32	rua	150	23	passar

151	23	podemos	191	18	governo
152	23	respeito	192	18	ladrões
153	23	seja	193	18	oque
154	22	ajuda	194	18	pagar
155	22	falar	195	18	sim
156	22	lugar	196	18	so
157	22	nunca	197	18	tipo
158	22	país	198	17	agora
159	22	primeiro	199	17	ate
160	22	pro	200	17	carro
161	22	ver	201	17	deles
162	22	vez	202	17	está
163	21	foram	203	17	história
164	21	gente	204	17	ia
165	21	nosso	205	17	legal
166	21	pela	206	17	meio
167	21	suas	207	17	melhorar
168	21	tenho	208	17	qualquer
169	21	todas	209	17	ruim
170	20	logo	210	17	slot
171	20	mapa	211	17	vamos
172	20	muitos	212	17	vou
173	20	poder	213	16	amigo
174	20	quem	214	16	amor
175	20	sejam	215	16	dar
176	19	agente	216	16	dois
177	19	ainda	217	16	feliz
178	19	boa	218	16	fosse
179	19	esta	219	16	futebol
180	19	frente	220	16	meninos
181	19	goonies	221	16	num
182	19	gosto	222	16	triste
183	19	novo	223	15	essa
184	19	segurança	224	15	fui
185	19	vão	225	15	grande
186	18	algumas	226	15	ja
187	18	começa	227	15	novos
188	18	dentro	228	15	papel
189	18	estavam	229	15	parte
190	18	fim	230	15	poluição



231	15	somos	271	12	casas
232	14	ai	272	12	caso
233	14	alegre	273	12	chegar
234	14	aparece	274	12	chão
235	14	apenas	275	12	começar
236	14	atores	276	12	corrupção
237	14	ferias	277	12	ensino
238	14	gostaria	278	12	essas
239	14	homens	279	12	fazendo
240	14	jogar	280	12	historia
241	14	la	281	12	mauro
242	14	legais	282	12	mesma
243	14	medo	283	12	micky
244	14	outra	284	12	opinião
245	14	parem	285	12	possam
246	14	preconceito	286	12	praia
247	14	seriam	287	12	saber
248	14	sonho	288	12	sendo
249	14	sonhos	289	12	sou
250	14	violência	290	12	talvez
251	13	aprender	291	12	tivesse
252	13	colegas	292	12	trabalho
253	13	conseguem	293	12	tão
254	13	contra	294	12	verdade
255	13	fica	295	12	volta
256	13	férias	296	11	amiga
257	13	mal	297	11	animais
258	13	melhores	298	11	baixo
259	13	minhas	299	11	cara
260	13	momentos	300	11	causa
261	13	nossos	301	11	cedo
262	13	outro	302	11	começam
263	13	poderia	303	11	comida
264	13	policiais	304	11	conseguir
265	13	realmente	305	11	devemos
266	13	tanto	306	11	entre
267	13	teve	307	11	espero
268	13	umas	308	11	faz
269	13	viver	309	11	menino
270	12	acham	310	11	mickey

311	11	mulheres	351	9	alguma
312	11	navio	352	9	alguns
313	11	pegar	353	9	bastante
314	11	porto	354	9	busca
315	11	qualidade	355	9	conhecer
316	11	real	356	9	dificuldades
317	11	semana	357	9	encontram
318	11	teria	358	9	enquanto
319	11	toda	359	9	faria
320	11	violencia	360	9	fazem
321	10	alexandre	361	9	ficou
322	10	algo	362	9	fiquei
323	10	amigas	363	9	ganha
324	10	antes	364	9	garotos
325	10	boas	365	9	hora
326	10	destacaria	366	9	igualdade
327	10	deu	367	9	mel
328	10	disse	368	9	nesse
329	10	dormir	369	9	nova
330	10	fala	370	9	passou
331	10	felizes	371	9	pelas
332	10	ficam	372	9	postos
333	10	gostei	373	9	sala
334	10	iguais	374	9	salário
335	10	lado	375	9	te
336	10	mostra	376	9	uns
337	10	nome	377	9	varias
338	10	novas	378	8	akira
339	10	numa	379	8	alguem
340	10	oportunidade	380	8	assaltos
341	10	povo	381	8	assinar
342	10	precisa	382	8	avó
343	10	principalmente	383	8	bala
344	10	quer	384	8	bocão
345	10	ruins	385	8	bola
346	10	salario	386	8	caolho
347	10	tu	387	8	certeza
348	10	usar	388	8	certo
349	10	valor	389	8	chega
350	9	aconteceu	390	8	cheio

391	8	colegio	431	7	achei
392	8	começou	432	7	acordar
393	8	conseguiu	433	7	afinal
394	8	contar	434	7	ajudando
395	8	continuar	435	7	alguém
396	8	cor	436	7	alunos
397	8	crimes	437	7	andar
398	8	cuidar	438	7	arrumar
399	8	daí	439	7	big
400	8	disso	440	7	chegou
401	8	estou	441	7	dinda
402	8	i	442	7	divertir
403	8	irmã	443	7	dão
404	8	irmãos	444	7	eliminaria
405	8	juntos	445	7	empregos
406	8	manhã	446	7	encontra
407	8	motivo	447	7	encontraram
408	8	mudaria	448	7	estudo
409	8	normal	449	7	existe
410	8	nossa	450	7	falando
411	8	parar	451	7	falta
412	8	passado	452	7	família
413	8	paz	453	7	ferreira
414	8	pensar	454	7	filho
415	8	pobre	455	7	filhos
416	8	podia	456	7	fugir
417	8	população	457	7	ganhar
418	8	possa	458	7	garoto
419	8	problemas	459	7	homem
420	8	protagonista	460	7	humanos
421	8	quase	461	7	inteiro
422	8	roubo	462	7	jeito
423	8	saúde	463	7	jogador
424	8	texto	464	7	julgar
425	8	time	465	7	junto
426	8	tinham	466	7	ladrão
427	8	viu	467	7	loja
428	8	vó	468	7	michael
429	7	acha	469	7	mik
430	7	acharam	470	7	nasci

471	7	natureza	511	6	deixar
472	7	negro	512	6	desafios
473	7	ninguem	513	6	desse
474	7	nisso	514	6	deveria
475	7	parece	515	6	direito
476	7	partes	516	6	dizem
477	7	passa	517	6	drogas
478	7	pelos	518	6	durante
479	7	pequena	519	6	entram
480	7	planeta	520	6	entrar
481	7	policia	521	6	eo
482	7	politicos	522	6	escapar
483	7	posso	523	6	estar
484	7	precisam	524	6	exemplo
485	7	primeira	525	6	familias
486	7	principal	526	6	fez
487	7	professor	527	6	fiz
488	7	rotina	528	6	fomos
489	7	ta	529	6	fora
490	7	tendo	530	6	gabriel
491	7	tirar	531	6	gosta
492	7	trabalhar	532	6	governadores
493	7	vejo	533	6	guerra
494	7	vem	534	6	igual
495	6	acabam	535	6	indo
496	6	acontecer	536	6	infancia
497	6	acreditar	537	6	isto
498	6	adrison	538	6	lareira
499	6	após	539	6	lição
500	6	aqueles	540	6	local
501	6	brincar	541	6	lugares
502	6	centro	542	6	luta
503	6	chamado	543	6	lutar
504	6	chegando	544	6	maik
505	6	cidade	545	6	maior
506	6	cima	546	6	maioria
507	6	começo	547	6	materiais
508	6	comigo	548	6	matheus
509	6	cristiano	549	6	melhore
510	6	dai	550	6	meses

551	6	momento	591	5	assaltada
552	6	muda	592	5	atenção
553	6	nossas	593	5	atras
554	6	pensando	594	5	atrás
555	6	penso	595	5	aventuras
556	6	perigo	596	5	aí
557	6	poderem	597	5	basquete
558	6	possível	598	5	brasileiro
559	6	precisamos	599	5	cadeia
560	6	prefeito	600	5	cadeias
561	6	prefeitura	601	5	campo
562	6	prisioneiros	602	5	cartas
563	6	prisão	603	5	caverna
564	6	queremos	604	5	chance
565	6	raça	605	5	chegam
566	6	respeitar	606	5	chegaram
567	6	rio	607	5	cinema
568	6	saco	608	5	comprar
569	6	salvar	609	5	condições
570	6	saude	610	5	consegue
571	6	sei	611	5	conseguiram
572	6	sermos	612	5	dado
573	6	serviço	613	5	dando
574	6	tarde	614	5	deixa
575	6	terra	615	5	desistir
576	6	torneio	616	5	diferente
577	6	ventiladores	617	5	divertido
578	6	è	618	5	dizendo
579	5	acaba	619	5	dizer
580	5	acabar	620	5	embora
581	5	achar	621	5	emprego
582	5	achou	622	5	eramos
583	5	achão	623	5	esperando
584	5	acidente	624	5	esporte
585	5	acontece	625	5	facil
586	5	americano	626	5	falaria
587	5	amo	627	5	fogo
588	5	apos	628	5	for
589	5	aquelas	629	5	grupo
590	5	arthur	630	5	hipoteca

631	5	horas	671	5	realizar
632	5	hospitais	672	5	recebem
633	5	hospital	673	5	reencontrar
634	5	humildes	674	5	rico
635	5	há	675	5	ronaldo
636	5	importa	676	5	roteiro
637	5	jogo	677	5	roupa
638	5	ladroes	678	5	roupas
639	5	limpo	679	5	sabe
640	5	lindo	680	5	sabia
641	5	matando	681	5	scott
642	5	matar	682	5	seguro
643	5	menas	683	5	serem
644	5	mike	684	5	tentacion
645	5	morar	685	5	teriam
646	5	morrendo	686	5	terminar
647	5	mude	687	5	tive
648	5	mulher	688	5	to
649	5	más	689	5	tomar
650	5	naõ	690	5	tv
651	5	nenhum	691	5	varios
652	5	nessa	692	5	velho
653	5	noite	693	5	vender
654	5	novamente	694	5	vergonha
655	5	ouro	695	5	vi
656	5	pare	696	5	vilões
657	5	passagem	697	5	vê
658	5	passam	698	5	xxx
659	5	pedras	699	4	abandonado
660	5	personagens	700	4	abreu
661	5	perto	701	4	acontecem
662	5	poluir	702	4	acordo
663	5	porta	703	4	adotado
664	5	presidente	704	4	adultos
665	5	presos	705	4	alegria
666	5	pros	706	4	algum
667	5	proximo	707	4	ali
668	5	pudesse	708	4	alimentos
669	5	públicas	709	4	além
670	5	racismo	710	4	amazonia

711	4	amazônia	751	4	condição
712	4	ambiente	752	4	consegui
713	4	amizade	753	4	contarei
714	4	antigo	754	4	contas
715	4	aos	755	4	conversar
716	4	aparecem	756	4	corrida
717	4	aprendemos	757	4	corruptos
718	4	aprendi	758	4	cortar
719	4	araujo	759	4	creches
720	4	armadilhas	760	4	criança
721	4	arvores	761	4	daria
722	4	atividades	762	4	denovo
723	4	aumentar	763	4	deus
724	4	baratas	764	4	deve
725	4	barco	765	4	diferença
726	4	bicicleta	766	4	difícil
727	4	bons	767	4	dizia
728	4	botar	768	4	dr
729	4	brendha	769	4	drama
730	4	bruxo	770	4	duas
731	4	bullyng	771	4	duro
732	4	buracos	772	4	empregada
733	4	cachorros	773	4	encontrar
734	4	cai	774	4	engraçado
735	4	cama	775	4	errado
736	4	caminho	776	4	espera
737	4	camiseta	777	4	estamos
738	4	caça	778	4	faculdade
739	4	celular	779	4	falei
740	4	chamada	780	4	fazia
741	4	chato	781	4	faço
742	4	chego	782	4	feita
743	4	cirurgia	783	4	feito
744	4	claro	784	4	felicidade
745	4	colocar	785	4	ficaram
746	4	colocaria	786	4	fico
747	4	comedia	787	4	fique
748	4	comendo	788	4	fizer
749	4	comer	789	4	fizeram
750	4	comum	790	4	forma

791	4	ganham	831	4	pump
792	4	idade	832	4	quanto
793	4	importante	833	4	quarto
794	4	independente	834	4	queimadas
795	4	infância	835	4	queira
796	4	inteligente	836	4	querer
797	4	iriam	837	4	realidade
798	4	jennifer	838	4	recreio
799	4	jovens	839	4	religião
800	4	justo	840	4	respeitados
801	4	levar	841	4	resto
802	4	lil	842	4	rodar
803	4	lixos	843	4	roubos
804	4	lições	844	4	sabem
805	4	medicos	845	4	sabemos
806	4	mendigo	846	4	saia
807	4	menina	847	4	saiu
808	4	moradores	848	4	salarios
809	4	morre	849	4	saudade
810	4	mortes	850	4	seguinte
811	4	mês	851	4	seres
812	4	nao	852	4	situação
813	4	nasceu	853	4	situações
814	4	nela	854	4	tentando
815	4	né	855	4	tentar
816	4	olá	856	4	terreno
817	4	paradas	857	4	tirando
818	4	parentes	858	4	tirei
819	4	passaram	859	4	tiveram
820	4	pediu	860	4	trabalhos
821	4	pegou	861	4	tristes
822	4	perdida	862	4	três
823	4	perfeito	863	4	verão
824	4	poderiam	864	4	via
825	4	pontos	865	4	viajar
826	4	porão	866	4	várias
827	4	pras	867	4	vários
828	4	prejudicar	868	4	°
829	4	problema	869	4	à
830	4	publicas	870	4	árvores



871	3	aconteceram	911	3	cancer
872	3	aconteceria	912	3	cansado
873	3	adianta	913	3	caras
874	3	agradeço	914	3	carater
875	3	ajudam	915	3	carros
876	3	ajude	916	3	chamo
877	3	alã	917	3	chata
878	3	aonde	918	3	idades
879	3	aparecendo	919	3	cinco
880	3	apesar	920	3	comandante
881	3	apoio	921	3	começão
882	3	aque	922	3	confiança
883	3	aquela	923	3	conseguimos
884	3	aquele	924	3	conta
885	3	aqui	925	3	contrato
886	3	aquilo	926	3	corpo
887	3	ar	927	3	correr
888	3	armadilha	928	3	criar
889	3	arume	929	3	cristo
890	3	assalto	930	3	cuide
891	3	atrizes	931	3	cuidem
892	3	aventura	932	3	culpa
893	3	ação	933	3	dardo
894	3	bagunça	934	3	dei
895	3	banho	935	3	deixando
896	3	baratos	936	3	deixou
897	3	barulho	937	3	dela
898	3	basta	938	3	delas
899	3	batia	939	3	demos
900	3	beleza	940	3	denzel
901	3	bolitas	941	3	desde
902	3	botam	942	3	desejo
903	3	botou	943	3	desista
904	3	brad	944	3	deveriam
905	3	brigas	945	3	devolta
906	3	brincava	946	3	dificuldade
907	3	cabeça	947	3	dignidade
908	3	caixa	948	3	diretor
909	3	caminhar	949	3	disoro
910	3	canadá	950	3	empatia

951	3	en	991	3	inporta
952	3	enfim	992	3	investir
953	3	enorme	993	3	irei
954	3	entra	994	3	iriamos
955	3	entraram	995	3	joga
956	3	eram	996	3	jogando
957	3	escolha	997	3	joias
958	3	escrevia	998	3	lanche
959	3	esgoto	999	3	lar
960	3	esses	1000	3	lei
961	3	estados	1001	3	levou
962	3	este	1002	3	licão
963	3	estranho	1003	3	liga
964	3	estrupe	1004	3	lojas
965	3	estudos	1005	3	lutando
966	3	exista	1006	3	mae
967	3	existiria	1007	3	manha
968	3	expressar	1008	3	mar
969	3	fabio	1009	3	mares
970	3	falam	1010	3	materias
971	3	falou	1011	3	mateus
972	3	faltar	1012	3	maus
973	3	familiares	1013	3	mcm
974	3	fazemos	1014	3	melhorem
975	3	festa	1015	3	meninas
976	3	filmes	1016	3	messi
977	3	fiquem	1017	3	mestre
978	3	fisica	1018	3	mina
979	3	fome	1019	3	modo
980	3	formar	1020	3	moro
981	3	forte	1021	3	morte
982	3	fossem	1022	3	mostrar
983	3	fugindo	1023	3	mostro
984	3	golfe	1024	3	mudou
985	3	greves	1025	3	museu
986	3	haja	1026	3	mão
987	3	humano	1027	3	música
988	3	ilusão	1028	3	namorado
989	3	inicio	1029	3	nele
990	3	inimigos	1030	3	neymar

1031	3	normalmente	1071	3	próximo
1032	3	notas	1072	3	publica
1033	3	objetivo	1073	3	publico
1034	3	objetivos	1074	3	publicos
1035	3	obras	1075	3	pé
1036	3	ola	1076	3	qual
1037	3	opnião	1077	3	querem
1038	3	palavra	1078	3	querendo
1039	3	papeis	1079	3	queriam
1040	3	passando	1080	3	rei
1041	3	pegam	1081	3	relembrar
1042	3	pegaram	1082	3	resolvem
1043	3	pensa	1083	3	respeitassem
1044	3	pensam	1084	3	resumindo
1045	3	pensava	1085	3	rever
1046	3	perdido	1086	3	risco
1047	3	pergunta	1087	3	roubando
1048	3	perguntei	1088	3	roubar
1049	3	pessos	1089	3	ruin
1050	3	pessoas	1090	3	sai
1051	3	pio	1091	3	saio
1052	3	pirata	1092	3	sairam
1053	3	plantas	1093	3	salvando
1054	3	pobres	1094	3	saquinho
1055	3	podre	1095	3	saíram
1056	3	ponto	1096	3	secreta
1057	3	pouquinho	1097	3	seguida
1058	3	preciosas	1098	3	seguiram
1059	3	precisando	1099	3	segura
1060	3	prefeitos	1100	3	sentar
1061	3	preocupar	1101	3	sequestro
1062	3	presente	1102	3	seriamos
1063	3	presidentes	1103	3	serie
1064	3	preços	1104	3	servir
1065	3	prima	1105	3	sexual
1066	3	primeiros	1106	3	sheron
1067	3	procurar	1107	3	si
1068	3	professoras	1108	3	sofrendo
1069	3	profissão	1109	3	sorte
1070	3	pré	1110	3	superação

1111	3	sustentar	1151	3	ótimo
1112	3	tais	1152	2	abandono
1113	3	tenhamos	1153	2	abraçei
1114	3	terem	1154	2	abria
1115	3	tesouros	1155	2	abrir
1116	3	tinhamos	1156	2	absurdo
1117	3	tipos	1157	2	abusar
1118	3	tiro	1158	2	acabou
1119	3	tiver	1159	2	aconteça
1120	3	tiverem	1160	2	acorda
1121	3	tivessem	1161	2	acreditava
1122	3	trabalhadores	1162	2	adolescentes
1123	3	tranquila	1163	2	adolescentes
1124	3	travis	1164	2	adota
1125	3	tristeza	1165	2	adotou
1126	3	tá	1166	2	adulta
1127	3	támbem	1167	2	agito
1128	3	ultima	1168	2	agradecer
1129	3	unidos	1169	2	agrações
1130	3	vagas	1170	2	ah
1131	3	valorizadas	1171	2	ajudem
1132	3	vam	1172	2	ajudo
1133	3	ve	1173	2	ajudou
1134	3	vidas	1174	2	alcançar
1135	3	vim	1175	2	almoçar
1136	3	vindo	1176	2	altas
1137	3	vir	1177	2	altos
1138	3	viram	1178	2	amar
1139	3	virar	1179	2	ameaçado
1140	3	visão	1180	2	amizades
1141	3	vivem	1181	2	americano
1142	3	viveria	1182	2	anciedade
1143	3	vivia	1183	2	anda
1144	3	vocês	1184	2	angelina
1145	3	voltamos	1185	2	animal
1146	3	vontade	1186	2	aparecia
1147	3	vô	1187	2	aparessem
1148	3	willy	1188	2	aprenderam
1149	3	á	1189	2	aprendeu
1150	3	água	1190	2	aproveitando

1191	2	aproveitar	1231	2	caneta
1192	2	armamentos	1232	2	características
1193	2	armarios	1233	2	carne
1194	2	armas	1234	2	caro
1195	2	arrumam	1235	2	carta
1196	2	artes	1236	2	cativeiro
1197	2	asaltados	1237	2	causar
1198	2	assaltadas	1238	2	cena
1199	2	assaltado	1239	2	chamava
1200	2	assaltantes	1240	2	chegamos
1201	2	assedida	1241	2	chegava
1202	2	assistindo	1242	2	cheguei
1203	2	assuntos	1243	2	cheia
1204	2	atitudes	1244	2	chorei
1205	2	ator	1245	2	clichê
1206	2	atraz	1246	2	clube
1207	2	avança	1247	2	coco
1208	2	avião	1248	2	colaborar
1209	2	baixos	1249	2	colegios
1210	2	banda	1250	2	coloca
1211	2	banheiro	1251	2	colorido
1212	2	banheiros	1252	2	colégio
1213	2	barato	1253	2	combater
1214	2	batalha	1254	2	comecei
1215	2	batalhão	1255	2	comemorar
1216	2	bk	1256	2	comemos
1217	2	bolita	1257	2	começaram
1218	2	bonzinho	1258	2	começaríamos
1219	2	bota	1259	2	completamente
1220	2	botaram	1260	2	completo
1221	2	branca	1261	2	comprou
1222	2	brien	1262	2	conclusão
1223	2	briga	1263	2	confiar
1224	2	brincavamos	1264	2	conhecemos
1225	2	bueiros	1265	2	conhecendo
1226	2	bullygn	1266	2	conheço
1227	2	cabelo	1267	2	conquista
1228	2	cadeiras	1268	2	consequia
1229	2	cadera	1269	2	consertem
1230	2	calçadas	1270	2	construimos

1271	2	construir	1311	2	destaquei
1272	2	contava	1312	2	devem
1273	2	contei	1313	2	deveríamos
1274	2	conteúdo	1314	2	deêm
1275	2	continua	1315	2	diamantes
1276	2	continuam	1316	2	diferenaa
1277	2	conversamos	1317	2	diferentes
1278	2	correndo	1318	2	diferenças
1279	2	corrupto	1319	2	difícil
1280	2	corupção	1320	2	diminuir
1281	2	cover	1321	2	direitos
1282	2	credito	1322	2	diversas
1283	2	creio	1323	2	divez
1284	2	crianças	1324	2	diz
1285	2	crime	1325	2	docas
1286	2	crise	1326	2	documento
1287	2	crush	1327	2	doente
1288	2	cuidamos	1328	2	dona
1289	2	cura	1329	2	dylan
1290	2	dança	1330	2	educada
1291	2	dançarinas	1331	2	educados
1292	2	dançarinos	1332	2	eliminar
1293	2	daqui	1333	2	empresas
1294	2	dava	1334	2	encontrou
1295	2	decide	1335	2	engraçadas
1296	2	defeitos	1336	2	ensina
1297	2	deformado	1337	2	ensinada
1298	2	deixarem	1338	2	entender
1299	2	depende	1339	2	entramos
1300	2	depressão	1340	2	entrando
1301	2	desabafar	1341	2	entregar
1302	2	descanso	1342	2	epoca
1303	2	desce	1343	2	equipe
1304	2	descer	1344	2	erradas
1305	2	desceram	1345	2	escolher
1306	2	desempregadas	1346	2	escolheria
1307	2	desigualdade	1347	2	escolinha
1308	2	desmoronando	1348	2	esconde
1309	2	desmoronar	1349	2	esconder
1310	2	dessistir	1350	2	escrever

1351	2	esperam	1391	2	ganhamos
1352	2	esperar	1392	2	ganhando
1353	2	esportes	1393	2	ganhei
1354	2	estado	1394	2	ganhou
1355	2	estariam	1395	2	garrafa
1356	2	estas	1396	2	geladeira
1357	2	estavamos	1397	2	gonies
1358	2	estreia	1398	2	gordinho
1359	2	estrupadas	1399	2	gostava
1360	2	eternamente	1400	2	governantes
1361	2	ex	1401	2	graças
1362	2	experiencias	1402	2	grupos
1363	2	experiência	1403	2	guaiba
1364	2	faca	1404	2	guerreira
1365	2	falava	1405	2	havia
1366	2	famosas	1406	2	helen
1367	2	famosos	1407	2	historias
1368	2	fassão	1408	2	histórias
1369	2	fato	1409	2	iam
1370	2	fazerem	1410	2	ideia
1371	2	ficado	1411	2	imaginação
1372	2	ficamos	1412	2	imediatamente
1373	2	ficando	1413	2	importantes
1374	2	ficarem	1414	2	importo
1375	2	ficariam	1415	2	impostos
1376	2	filha	1416	2	infantil
1377	2	filhotes	1417	2	infelizmente
1378	2	floresta	1418	2	infelizmente
1379	2	florestas	1419	2	inglês
1380	2	formados	1420	2	inocente
1381	2	força	1421	2	inocentes
1382	2	frio	1422	2	inteira
1383	2	fuga	1423	2	internacional
1384	2	fugiram	1424	2	interpretado
1385	2	funciona	1425	2	inventar
1386	2	funcionem	1426	2	invenção
1387	2	furtos	1427	2	investido
1388	2	futuramente	1428	2	irem
1389	2	fáceis	1429	2	irão
1390	2	fácil	1430	2	jackie

1431	2	janeiro	1471	2	melhorias
1432	2	jantar	1472	2	melissa
1433	2	jesus	1473	2	mendigos
1434	2	jogamos	1474	2	mentir
1435	2	jogavamos	1475	2	merece
1436	2	jogou	1476	2	mesmas
1437	2	joguem	1477	2	mesmos
1438	2	jornada	1478	2	meteoro
1439	2	jugar	1479	2	mico
1440	2	julgassem	1480	2	milhão
1441	2	lealdade	1481	2	militar
1442	2	lembro	1482	2	minoria
1443	2	ler	1483	2	mistérios
1444	2	leva	1484	2	mochila
1445	2	liberar	1485	2	montar
1446	2	lien	1486	2	montro
1447	2	ligam	1487	2	mora
1448	2	limpar	1488	2	morando
1449	2	linha	1489	2	morei
1450	2	livre	1490	2	mortas
1451	2	livres	1491	2	motivos
1452	2	livros	1492	2	moça
1453	2	lixeiros	1493	2	mudando
1454	2	lo	1494	2	mudança
1455	2	los	1495	2	mudaram
1456	2	louco	1496	2	mudasse
1457	2	luzes	1497	2	mudei
1458	2	machismo	1498	2	naquela
1459	2	madrugada	1499	2	nasce
1460	2	maravilhosa	1500	2	nascer
1461	2	maravilhosos	1501	2	nascimento
1462	2	marca	1502	2	natal
1463	2	marciais	1503	2	necessário
1464	2	martelo	1504	2	negra
1465	2	material	1505	2	neles
1466	2	matue	1506	2	nenhuma
1467	2	matérias	1507	2	nesses
1468	2	mau	1508	2	nesta
1469	2	mecho	1509	2	neste
1470	2	medio	1510	2	ninguém



1511	2	ninjas	1551	2	pessoal
1512	2	nivel	1552	2	pia
1513	2	nota	1553	2	pires
1514	2	notei	1554	2	piscina
1515	2	odeio	1555	2	play
1516	2	olha	1556	2	podesse
1517	2	olhando	1557	2	policial
1518	2	olhasse	1558	2	política
1519	2	olhos	1559	2	políticos
1520	2	onibus	1560	2	posição
1521	2	ouvir	1561	2	poucas
1522	2	pacote	1562	2	praias
1523	2	pagasem	1563	2	pragas
1524	2	países	1564	2	praça
1525	2	param	1565	2	precarias
1526	2	parassem	1566	2	precisaram
1527	2	parecidos	1567	2	precisosas
1528	2	paredes	1568	2	prego
1529	2	participantes	1569	2	preguiça
1530	2	particular	1570	2	preguiçosa
1531	2	particulares	1571	2	prejudica
1532	2	partimos	1572	2	premio
1533	2	passarem	1573	2	prende
1534	2	países	1574	2	prender
1535	2	pede	1575	2	prestes
1536	2	pedir	1576	2	pretende
1537	2	pegando	1577	2	preço
1538	2	pegue	1578	2	principais
1539	2	pena	1579	2	privacidade
1540	2	pensamento	1580	2	procura
1541	2	pensamos	1581	2	produções
1542	2	pensavam	1582	2	profissões
1543	2	pensem	1583	2	profissão
1544	2	pequeno	1584	2	proprio
1545	2	perder	1585	2	protagonizado
1546	2	perdidas	1586	2	proteja
1547	2	pergunto	1587	2	protesto
1548	2	perseguição	1588	2	prova
1549	2	personagem	1589	2	provavelmente
1550	2	pessoais	1590	2	prédios

1591	2	públicos	1631	2	seguranças
1592	2	quatro	1632	2	semanas
1593	2	quausa	1633	2	senhor
1594	2	queimar	1634	2	sentir
1595	2	queelas	1635	2	separar
1596	2	quele	1636	2	sequestros
1597	2	quinta	1637	2	sera
1598	2	quis	1638	2	significa
1599	2	quiser	1639	2	simples
1600	2	rainha	1640	2	sinto
1601	2	rampas	1641	2	sloott
1602	2	rascismo	1642	2	sobreviver
1603	2	recebendo	1643	2	sociedade
1604	2	recebo	1644	2	sofrer
1605	2	reclamar	1645	2	sofrido
1606	2	recomendo	1646	2	sol
1607	2	recomeçar	1647	2	solução
1608	2	refugiados	1648	2	sonhava
1609	2	remedios	1649	2	sotan
1610	2	remédio	1650	2	sotão
1611	2	repita	1651	2	subo
1612	2	resgatar	1652	2	suficiente
1613	2	resolveram	1653	2	suja
1614	2	ricos	1654	2	sujeira
1615	2	rima	1655	2	sujo
1616	2	romanticas	1656	2	sul
1617	2	roubem	1657	2	sótão
1618	2	rs	1658	2	tal
1619	2	saben	1659	2	tambèm
1620	2	sabendo	1660	2	tanta
1621	2	sabiam	1661	2	tarefas
1622	2	saiba	1662	2	tata
1623	2	saida	1663	2	tava
1624	2	sairem	1664	2	taís
1625	2	sairia	1665	2	tecnologia
1626	2	salas	1666	2	tenta
1627	2	salva	1667	2	tentam
1628	2	sedo	1668	2	tentou
1629	2	seguindo	1669	2	tera
1630	2	seguir	1670	2	terminem

1671	2	termino	1711	2	voltei
1672	2	titulos	1712	2	volto
1673	2	tornar	1713	2	vá
1674	2	torneios	1714	2	willi
1675	2	totalmente	1715	1	abaixam
1676	2	trabalha	1716	1	abandonados
1677	2	trabalham	1717	1	abandone
1678	2	trafico	1718	1	abc
1679	2	trancou	1719	1	aberta
1680	2	traria	1720	1	abiental
1681	2	trazer	1721	1	abisurdo
1682	2	trevas	1722	1	abolir
1683	2	troca	1723	1	aborta
1684	2	trouxa	1724	1	abre
1685	2	tráfico	1725	1	abri
1686	2	turmas	1726	1	abrigos
1687	2	turísticos	1727	1	abrila
1688	2	ultimo	1728	1	abriu
1689	2	unida	1729	1	abusada
1690	2	unão	1730	1	abusadas
1691	2	valorizada	1731	1	abusam
1692	2	valorizem	1732	1	abuso
1693	2	vasculhar	1733	1	acabamos
1694	2	vemos	1734	1	acabe
1695	2	venda	1735	1	acada
1696	2	vendo	1736	1	acaso
1697	2	ventilador	1737	1	aceitam
1698	2	verbas	1738	1	aceitarem
1699	2	verdadeira	1739	1	acendendo
1700	2	verdadeiras	1740	1	achando
1701	2	viagem	1741	1	acharia
1702	2	video	1742	1	achãm
1703	2	vilas	1743	1	acima
1704	2	visitar	1744	1	acomdesendo
1705	2	vivemos	1745	1	aconteca
1706	2	vivermos	1746	1	acontecendo
1707	2	vivos	1747	1	aconteceriam
1708	2	volei	1748	1	acontecido
1709	2	voltaram	1749	1	acontecimentos
1710	2	voltaria	1750	1	aconteçido

1751	1	acordado	1791	1	agregção
1752	1	acordando	1792	1	agricividade
1753	1	acordava	1793	1	agua
1754	1	acorrentado	1794	1	aguentando
1755	1	acostumar	1795	1	aguento
1756	1	acre	1796	1	air
1757	1	acredita	1797	1	aja
1758	1	acreditan	1798	1	ajeita
1759	1	acredite	1799	1	ajuase
1760	1	acredito	1800	1	ajudace
1761	1	acreditou	1801	1	ajudamos
1762	1	acumulado	1802	1	ajudarem
1763	1	acumulando	1803	1	ajudaria
1764	1	adaptando	1804	1	ajudavam
1765	1	adiantar	1805	1	ajudei
1766	1	admirado	1806	1	ajudá
1767	1	adimiram	1807	1	ajudão
1768	1	admiro	1808	1	ajuntando
1769	1	adoeceu	1809	1	ajustando
1770	1	adolescência	1810	1	alagadas
1771	1	adolescencia	1811	1	alan
1772	1	adolescente	1812	1	alberto
1773	1	adolescentes	1813	1	aldeia
1774	1	adolescência	1814	1	alegrar
1775	1	adolêscencia	1815	1	alem
1776	1	adoro	1816	1	alfalto
1777	1	adotada	1817	1	alguen
1778	1	adotados	1818	1	algul
1779	1	adotasem	1819	1	algumais
1780	1	adubar	1820	1	algun
1781	1	adversario	1821	1	alicerce
1782	1	aeroporto	1822	1	aliás
1783	1	aeste	1823	1	almentar
1784	1	afeganistão	1824	1	almocei
1785	1	afeta	1825	1	almoçava
1786	1	afunde	1826	1	almoço
1787	1	agir	1827	1	alopra
1788	1	agitada	1828	1	alquem
1789	1	agradavel	1829	1	alsatos
1790	1	agressões	1830	1	alta

1831	1	alto	1871	1	aparência
1832	1	altura	1872	1	aparesse
1833	1	alumínio	1873	1	apartir
1834	1	aléxia	1874	1	aparência
1835	1	ama	1875	1	ape
1836	1	amadas	1876	1	apegar
1837	1	amado	1877	1	apelido
1838	1	amadurecimento	1878	1	aponada
1839	1	amamos	1879	1	aposentadoria
1840	1	amanhã	1880	1	apouco
1841	1	amarela	1881	1	apredi
1842	1	amassem	1882	1	apredisagem
1843	1	amaveis	1883	1	apredizado
1844	1	amazonas	1884	1	aprenda
1845	1	amem	1885	1	aprendam
1846	1	amigável	1886	1	aprende
1847	1	amos	1887	1	aprenderem
1848	1	amvep	1888	1	aprendido
1849	1	anciosa	1889	1	aprendizado
1850	1	ancioso	1890	1	aprendizados
1851	1	andan	1891	1	apresenta
1852	1	andando	1892	1	apresentam
1853	1	andaram	1893	1	aprontando
1854	1	ando	1894	1	aproveitam
1855	1	angulos	1895	1	aproveite
1856	1	angustiado	1896	1	aproventam
1857	1	animado	1897	1	area
1858	1	aninho	1898	1	areia
1859	1	anoitecer	1899	1	arra
1860	1	ante	1900	1	arrecadaria
1861	1	antigamente	1901	1	arrecem
1862	1	antigos	1902	1	arresem
1863	1	antis	1903	1	arriscando
1864	1	aode	1904	1	arrisco
1865	1	apagada	1905	1	arrois
1866	1	apagando	1906	1	arruma
1867	1	apaixonada	1907	1	arrumace
1868	1	apaixonaria	1908	1	arrumem
1869	1	aparecerem	1909	1	artista
1870	1	aparelhos	1910	1	artur

1911	1	arvore	1951	1	atividade
1912	1	arvóres	1952	1	ativão
1913	1	asalta	1953	1	atrapalham
1914	1	asaltam	1954	1	atrasado
1915	1	asautos	1955	1	atrasar
1916	1	aspectos	1956	1	atravessei
1917	1	aspirantes	1957	1	atraído
1918	1	assaltados	1958	1	atris
1919	1	assaltante	1959	1	atriz
1920	1	assaltar	1960	1	atuais
1921	1	assaltaria	1961	1	atualmente
1922	1	assando	1962	1	atuar
1923	1	assasino	1963	1	atura
1924	1	assassinato	1964	1	aturar
1925	1	assato	1965	1	aumentacem
1926	1	assaudado	1966	1	aumentando
1927	1	assautados	1967	1	aunemtamdo
1928	1	assedeadas	1968	1	auto
1929	1	assidade	1969	1	avançado
1930	1	assina	1970	1	averdade
1931	1	assinando	1971	1	avisar
1932	1	assistir	1972	1	avisou
1933	1	assombrado	1973	1	avós
1934	1	assustada	1974	1	avô
1935	1	assustado	1975	1	azar
1936	1	assustador	1976	1	ações
1937	1	assustados	1977	1	b
1938	1	assustar	1978	1	baah
1939	1	assédio	1979	1	bag
1940	1	asteriscos	1980	1	bagunsava
1941	1	astoria	1981	1	bagunçando
1942	1	asvezes	1982	1	bahia
1943	1	atearam	1983	1	bairo
1944	1	atenciosos	1984	1	bairro
1945	1	atenda	1985	1	baixa
1946	1	atendimento	1986	1	baleado
1947	1	atirado	1987	1	banais
1948	1	atirar	1988	1	banco
1949	1	atitude	1989	1	bandido
1950	1	ativam	1990	1	barradas

1991	1	barranco	2031	1	bonita
1992	1	barriga	2032	1	bonitas
1993	1	barulhos	2033	1	bonito
1994	1	base	2034	1	bonitos
1995	1	baseado	2035	1	borrado
1996	1	baseia	2036	1	botado
1997	1	basil	2037	1	botando
1998	1	basquetibol	2038	1	botaria
1999	1	bastantes	2039	1	botem
2000	1	batalhadora	2040	1	botijãozinho
2001	1	batalhando	2041	1	br
2002	1	batalhava	2042	1	brabeza
2003	1	bate	2043	1	bral
2004	1	batendo	2044	1	brancas
2005	1	bater	2045	1	branco
2006	1	bateram	2046	1	brancos
2007	1	batman	2047	1	brasileiros
2008	1	bebedor	2048	1	brassil
2009	1	beber	2049	1	bravo
2010	1	bebia	2050	1	bravura
2011	1	bebê	2051	1	brigadas
2012	1	beira	2052	1	brigando
2013	1	bela	2053	1	brigar
2014	1	belanie	2054	1	brigava
2015	1	belém	2055	1	brincadeiras
2016	1	ben	2056	1	brincam
2017	1	beneficiar	2057	1	brincamos
2018	1	benmal	2058	1	brincavam
2019	1	besouro	2059	1	brinquedos
2020	1	biblia	2060	1	briquedos
2021	1	bicicletas	2061	1	brown
2022	1	bisbilhotar	2062	1	bruxa
2023	1	boero	2063	1	bruxas
2024	1	bokão	2064	1	bruxos
2025	1	bolinhas	2065	1	buliticos
2026	1	bolsa	2066	1	bulling
2027	1	bolsonaro	2067	1	bullying
2028	1	bolsos	2068	1	bullyny
2029	1	bomero	2069	1	bulyng
2030	1	bondosa	2070	1	bunito

2071	1	buraco	2111	1	cargo
2072	1	buscar	2112	1	carinha
2073	1	básico	2113	1	carinho
2074	1	bônus	2114	1	carlos
2075	1	cabri	2115	1	carnaval
2076	1	cachorrinhos	2116	1	carolina
2077	1	cachorro	2117	1	carregar
2078	1	cade	2118	1	cartinha
2079	1	cadeira	2119	1	casaco
2080	1	cadelinha	2120	1	casar
2081	1	caderantes	2121	1	casca
2082	1	cadernos	2122	1	casinha
2083	1	cadê	2123	1	casos
2084	1	café	2124	1	castelo
2085	1	caindo	2125	1	catadores
2086	1	caiu	2126	1	catar
2087	1	caldas	2127	1	cativeiros
2088	1	calma	2128	1	causadas
2089	1	calmaria	2129	1	causando
2090	1	calmo	2130	1	cauçadas
2091	1	camera	2131	1	cavalo
2092	1	cameron	2132	1	cavera
2093	1	caminhando	2133	1	caçar
2094	1	camisetas	2134	1	ceara
2095	1	campeão	2135	1	ceculo
2096	1	camponesas	2136	1	cegas
2097	1	camuflados	2137	1	ceia
2098	1	cansando	2138	1	cela
2099	1	cansar	2139	1	cenas
2100	1	cansativo	2140	1	centineo
2101	1	canteiros	2141	1	centir
2102	1	canto	2142	1	certa
2103	1	cantoras	2143	1	certesa
2104	1	capa	2144	1	certessa
2105	1	capacidade	2145	1	certos
2106	1	capazes	2146	1	ceurugia
2107	1	capazs	2147	1	cham
2108	1	capturado	2148	1	chama
2109	1	capturados	2149	1	chamar
2110	1	caramba	2150	1	chamaram



2151	1	chamei	2191	1	comercializados
2152	1	chan	2192	1	comesando
2153	1	chances	2193	1	comessar
2154	1	chatas	2194	1	cometem
2155	1	chatos	2195	1	começando
2156	1	chave	2196	1	começaria
2157	1	chaves	2197	1	começei
2158	1	chegasse	2198	1	comidas
2159	1	cheguem	2199	1	compartilharia
2160	1	cheias	2200	1	competencia
2161	1	cheiro	2201	1	competindo
2162	1	chigande	2202	1	completar
2163	1	chip	2203	1	completasse
2164	1	choram	2204	1	completei
2165	1	choraram	2205	1	complicada
2166	1	chorava	2206	1	complicado
2167	1	choro	2207	1	comportar
2168	1	chões	2208	1	compra
2169	1	cidadões	2209	1	compraram
2170	1	cientifica	2210	1	compreendido
2171	1	cirugia	2211	1	compreensivo
2172	1	ciumes	2212	1	compreensão
2173	1	ciências	2213	1	comprem
2174	1	claramente	2214	1	comprienção
2175	1	cobertor	2215	1	comunidade
2176	1	cobrando	2216	1	comunitario
2177	1	cobravam	2217	1	comédia
2178	1	cody	2218	1	conceguir
2179	1	colaboracem	2219	1	concidero
2180	1	colega	2220	1	concluir
2181	1	colheu	2221	1	concorda
2182	1	collins	2222	1	concorrer
2183	1	colo	2223	1	concreto
2184	1	colocasse	2224	1	condicionado
2185	1	coloquei	2225	1	condicoes
2186	1	colégios	2226	1	confian
2187	1	colégio	2227	1	confiantes
2188	1	comandava	2228	1	confio
2189	1	combinações	2229	1	confiou
2190	1	comecem	2230	1	conflitos

2231	1	confuso	2271	1	contratar
2232	1	confusão	2272	1	contro
2233	1	confusões	2273	1	contruam
2234	1	congelado	2274	1	conversa
2235	1	conhece	2275	1	conversaram
2236	1	conhecemos	2276	1	conversava
2237	1	conhecíamos	2277	1	convertendo
2238	1	conhecido	2278	1	convertessemos
2239	1	conhecidos	2279	1	conviver
2240	1	conher	2280	1	convocado
2241	1	conquistas	2281	1	coordenadas
2242	1	consciência	2282	1	copiar
2243	1	conseguindo	2283	1	copo
2244	1	conseguiremos	2284	1	corajos
2245	1	conseguiria	2285	1	coração
2246	1	conseguissem	2286	1	coração
2247	1	conseguiu	2287	1	corido
2248	1	conselho	2288	1	corpos
2249	1	conselhos	2289	1	corretas
2250	1	consentiu	2290	1	correu
2251	1	consequência	2291	1	corria
2252	1	conserta	2292	1	corriamos
2253	1	conserteza	2293	1	corridas
2254	1	consientisen	2294	1	corrupção
2255	1	consiga	2295	1	cortadas
2256	1	consigui	2296	1	corte
2257	1	constante	2297	1	cortenação
2258	1	construção	2298	1	costas
2259	1	consultar	2299	1	cozinheir
2260	1	consultas	2300	1	cramas
2261	1	consumo	2301	1	crece
2262	1	contando	2302	1	crescer
2263	1	contaremos	2303	1	creceu
2264	1	contaria	2304	1	cresci
2265	1	continuando	2305	1	crescido
2266	1	continuação	2306	1	crescimento
2267	1	continuei	2307	1	criados
2268	1	conto	2308	1	crialos
2269	1	contou	2309	1	criansas
2270	1	contratado	2310	1	criança

2311	1	criarão	2351	1	deficiente
2312	1	criasse	2352	1	defunto
2313	1	criava	2353	1	deis
2314	1	criminosos	2354	1	deiso
2315	1	criou	2355	1	deitada
2316	1	cristian	2356	1	deito
2317	1	cruéis	2357	1	deixam
2318	1	críticas	2358	1	deixaram
2319	1	crudadas	2359	1	deixaria
2320	1	cuidadas	2360	1	deixasse
2321	1	cuidarmos	2361	1	dem
2322	1	curiosidades	2362	1	demite
2323	1	curzar	2363	1	demolir
2324	1	curso	2364	1	demoraria
2325	1	curti	2365	1	demore
2326	1	curtindo	2366	1	demorou
2327	1	custa	2367	1	dengosas
2328	1	custas	2368	1	denis
2329	1	cãozinho	2369	1	denominado
2330	1	dada	2370	1	dentre
2331	1	dali	2371	1	deparam
2332	1	dam	2372	1	deparar
2333	1	dançar	2373	1	depararam
2334	1	dançaram	2374	1	dependendo
2335	1	daquele	2375	1	dependente
2336	1	darmos	2376	1	depressão
2337	1	debaixo	2377	1	deputados
2338	1	debora	2378	1	der
2339	1	decaídas	2379	1	deram
2340	1	decente	2380	1	derepente
2341	1	decepções	2381	1	derrubando
2342	1	dechar	2382	1	derrubou
2343	1	decidida	2383	1	des
2344	1	decidido	2384	1	desarrumadas
2345	1	decidimos	2385	1	desastre
2346	1	decorrer	2386	1	descansando
2347	1	dedinho	2387	1	descoberto
2348	1	dedo	2388	1	descobrimo
2349	1	deficiente	2389	1	descobrirá
2350	1	definir	2390	1	descobriu

2391	1	desconhecida	2431	1	devessem
2392	1	desconhecido	2432	1	devia
2393	1	desculpa	2433	1	dezembro
2394	1	descurberta	2434	1	diante
2395	1	descurço	2435	1	difical
2396	1	desdo	2436	1	dificio
2397	1	desempenho	2437	1	dificultar
2398	1	desemprego	2438	1	difunto
2399	1	desempregos	2439	1	difíceis
2400	1	desenpregadas	2440	1	digamos
2401	1	desgraça	2441	1	digna
2402	1	desiste	2442	1	dignas
2403	1	desisti	2443	1	digno
2404	1	desistimos	2444	1	digo
2405	1	desistindo	2445	1	dilma
2406	1	desistiu	2446	1	diminuem
2407	1	desmatadas	2447	1	dinheiros
2408	1	desmatamentos	2448	1	dinhiro
2409	1	desmatação	2449	1	direto
2410	1	desnecessarios	2450	1	discarga
2411	1	despedida	2451	1	disfarçado
2412	1	despejados	2452	1	disney
2413	1	despejar	2453	1	disputa
2414	1	despejo	2454	1	disrespentam
2415	1	desperdisem	2455	1	distante
2416	1	despertar	2456	1	diversão
2417	1	despresar	2457	1	diverti
2418	1	desprezam	2458	1	divertia
2419	1	desrespeitasse	2459	1	divertida
2420	1	dessas	2460	1	divertimos
2421	1	destacando	2461	1	divertindo
2422	1	destaque	2462	1	divida
2423	1	destino	2463	1	dividia
2424	1	destrui	2464	1	doações
2425	1	destruida	2465	1	dobro
2426	1	destruir	2466	1	doces
2427	1	destruição	2467	1	doideras
2428	1	desvendaria	2468	1	dolare
2429	1	desçe	2469	1	dominarem
2430	1	devendo	2470	1	dono

2471	1	donos	2511	1	empenhou
2472	1	dor	2512	1	emporta
2473	1	dormimos	2513	1	emportem
2474	1	dormindo	2514	1	empregadas
2475	1	dotada	2515	1	envolvido
2476	1	dotor	2516	1	encamento
2477	1	dove	2517	1	encheram
2478	1	dramas	2518	1	enchi
2479	1	dramáticas	2519	1	enclusive
2480	1	droga	2520	1	encomoda
2481	1	dudu	2521	1	encomodando
2482	1	duenti	2522	1	encomodar
2483	1	dunada	2523	1	encontrado
2484	1	dupla	2524	1	encontrando
2485	1	durmo	2525	1	encontrariamos
2486	1	duvida	2526	1	encontro
2487	1	duvidar	2527	1	enforçar
2488	1	dzendo	2528	1	enfrentar
2489	1	dá	2529	1	enfrento
2490	1	dê	2530	1	engrasadas
2491	1	echem	2531	1	engravidam
2492	1	editor	2532	1	enjoando
2493	1	edosos	2533	1	enporta
2494	1	edu	2534	1	enportar
2495	1	eduardo	2535	1	enportariam
2496	1	educadas	2536	1	enpregada
2497	1	educar	2537	1	enredos
2498	1	educativo	2538	1	ensinado
2499	1	educação	2539	1	ensinam
2500	1	efeitos	2540	1	ensinaria
2501	1	efetuar	2541	1	entaõ
2502	1	egoismo	2542	1	entediados
2503	1	elaborado	2543	1	entendam
2504	1	eleição	2544	1	entendem
2505	1	eletrônicos	2545	1	entendi
2506	1	eliminei	2546	1	enteresado
2507	1	elimino	2547	1	interesse
2508	1	embarcariam	2548	1	enti
2509	1	emcarar	2549	1	entope
2510	1	emoções	2550	1	entrefista

2551	1	entregaram	2591	1	especiais
2552	1	entreguei	2592	1	especialmente
2553	1	entrevia	2593	1	especifico
2554	1	entrão	2594	1	espelhos
2555	1	enturmar	2595	1	esperamos
2556	1	envergonhado	2596	1	esperaram
2557	1	investido	2597	1	esperarmos
2558	1	enviadas	2598	1	esperavam
2559	1	enviado	2599	1	esperto
2560	1	enviados	2600	1	espiada
2561	1	enviou	2601	1	esqueceríamos
2562	1	envista	2602	1	esquecido
2563	1	epocas	2603	1	esqueleto
2564	1	equipamento	2604	1	esquerda
2565	1	equipamentos	2605	1	esqueção
2566	1	erika	2606	1	essenciais
2567	1	errados	2607	1	estadios
2568	1	errar	2608	1	estao
2569	1	errei	2609	1	estara
2570	1	erros	2610	1	estare
2571	1	esas	2611	1	estarem
2572	1	escapando	2612	1	estariamos
2573	1	escondem	2613	1	estatua
2574	1	esconderia	2614	1	estavan
2575	1	escondidas	2615	1	estavão
2576	1	escondidos	2616	1	esteve
2577	1	escorregou	2617	1	estima
2578	1	escrevem	2618	1	estiver
2579	1	escreveria	2619	1	estrada
2580	1	escrito	2620	1	estradas
2581	1	esculachadas	2621	1	estragaram
2582	1	escuta	2622	1	estrupos
2583	1	escutam	2623	1	estrupão
2584	1	escutar	2624	1	estrutura
2585	1	escute	2625	1	estruturas
2586	1	esdigies	2626	1	estudamos
2587	1	esforçar	2627	1	estudando
2588	1	esforço	2628	1	estudantes
2589	1	espanhol	2629	1	estudava
2590	1	espaço	2630	1	estudavamos

2631	1	estude	2671	1	falsidade
2632	1	estára	2672	1	falsos
2633	1	estância	2673	1	faltando
2634	1	et	2674	1	faltas
2635	1	etica	2675	1	falte
2636	1	eua	2676	1	falão
2637	1	europa	2677	1	familha
2638	1	evolua	2678	1	familiar
2639	1	exagero	2679	1	família
2640	1	exaltadas	2680	1	famoso
2641	1	exato	2681	1	famílias
2642	1	excepcional	2682	1	fasam
2643	1	excolheria	2683	1	fasso
2644	1	exculachou	2684	1	favor
2645	1	exemplos	2685	1	faze
2646	1	existia	2686	1	fazeria
2647	1	existiriam	2687	1	fazíamos
2648	1	existissem	2688	1	façam
2649	1	expectativa	2689	1	fechada
2650	1	explorada	2690	1	feci
2651	1	explorar	2691	1	fedorentos
2652	1	exploração	2692	1	feia
2653	1	explosão	2693	1	feio
2654	1	expulsando	2694	1	feios
2655	1	fabulosa	2695	1	feira
2656	1	facas	2696	1	feitiços
2657	1	facer	2697	1	felicidades
2658	1	faiz	2698	1	felizmente
2659	1	falado	2699	1	felton
2660	1	falamos	2700	1	feminicio
2661	1	falaram	2701	1	feministas
2662	1	falarem	2702	1	feridas
2663	1	falasse	2703	1	fernanda
2664	1	falavamos	2704	1	ferra
2665	1	falecer	2705	1	festinhas
2666	1	falecido	2706	1	ficamdo
2667	1	falesceu	2707	1	ficaremos
2668	1	falhas	2708	1	ficaria
2669	1	falo	2709	1	ficava
2670	1	falsa	2710	1	ficavam

2711	1	ficessem	2751	1	frezzer
2712	1	ficãm	2752	1	fria
2713	1	ficão	2753	1	frias
2714	1	ficção	2754	1	frorestas
2715	1	fidrido	2755	1	fucionarios
2716	1	fiennes	2756	1	fugi
2717	1	fila	2757	1	fuluro
2718	1	filhote	2758	1	fumar
2719	1	filipe	2759	1	fumaças
2720	1	filmagem	2760	1	fumcioma
2721	1	filmagens	2761	1	funcionam
2722	1	filmaço	2762	1	funcionando
2723	1	fims	2763	1	funcionar
2724	1	finais	2764	1	funcionarios
2725	1	financeiros	2765	1	funcionários
2726	1	fingindo	2766	1	fundamental
2727	1	fire	2767	1	fundo
2728	1	firmas	2768	1	fundos
2729	1	fixa	2769	1	furos
2730	1	fizemos	2770	1	futeboll
2731	1	fizesse	2771	1	futibol
2732	1	fizessem	2772	1	fuzil
2733	1	fliperama	2773	1	fuzileiros
2734	1	flores	2774	1	fuzis
2735	1	florianopolis	2775	1	fé
2736	1	folha	2776	1	game
2737	1	fonsiona	2777	1	ganancia
2738	1	footeboll	2778	1	ganharia
2739	1	foquei	2779	1	ganhasse
2740	1	forem	2780	1	ganhassem
2741	1	formata	2781	1	garantindo
2742	1	formatura	2782	1	garanto
2743	1	formaturas	2783	1	garimpeiros
2744	1	formei	2784	1	gasolina
2745	1	formigueiro	2785	1	gasometro
2746	1	fortes	2786	1	gastaria
2747	1	fox	2787	1	gaste
2748	1	fozem	2788	1	gays
2749	1	frases	2789	1	gelada
2750	1	free	2790	1	gemte



2791	1	generosos	2831	1	gás
2792	1	geografia	2832	1	gênero
2793	1	gerado	2833	1	haddid
2794	1	geral	2834	1	haver
2795	1	gigante	2835	1	haveria
2796	1	gigi	2836	1	herois
2797	1	gira	2837	1	heroismo
2798	1	gol	2838	1	higiene
2799	1	gols	2839	1	hilario
2800	1	gonis	2840	1	hogwart
2801	1	goon	2841	1	hollywod
2802	1	goonie	2842	1	homeis
2803	1	goonis	2843	1	homenage
2804	1	gorveno	2844	1	honestos
2805	1	gostam	2845	1	horario
2806	1	gostariam	2846	1	horriveis
2807	1	goste	2847	1	horrível
2808	1	gostem	2848	1	horário
2809	1	gostou	2849	1	hotel
2810	1	governado	2850	1	houve
2811	1	gratidão	2851	1	houvesse
2812	1	gravando	2852	1	hulk
2813	1	gravaria	2853	1	humildade
2814	1	gravida	2854	1	humilde
2815	1	graziele	2855	1	humilhado
2816	1	grazy	2856	1	humilhou
2817	1	graça	2857	1	humor
2818	1	greche	2858	1	humorados
2819	1	greve	2859	1	hóspitais
2820	1	grita	2860	1	ida
2821	1	grêmio	2861	1	identidade
2822	1	guando	2862	1	identifica
2823	1	guardam	2863	1	identificação
2824	1	guardaria	2864	1	idiotas
2825	1	guardava	2865	1	ido
2826	1	gude	2866	1	igieni
2827	1	guerras	2867	1	ignorantes
2828	1	guiba	2868	1	iludido
2829	1	guri	2869	1	imagens
2830	1	guria	2870	1	imaginando

2871	1	imaginar	2911	1	interiro
2872	1	imagine	2912	1	interpretada
2873	1	imaturidade	2913	1	interpretados
2874	1	impedindo	2914	1	interropções
2875	1	importancia	2915	1	intão
2876	1	importando	2916	1	invens
2877	1	importar	2917	1	inverno
2878	1	importava	2918	1	inves
2879	1	impossivel	2919	1	invista
2880	1	impossível	2920	1	invistam
2881	1	imposto	2921	1	início
2882	1	incendio	2922	1	iorgute
2883	1	incenções	2923	1	ipócritas
2884	1	incomum	2924	1	ira
2885	1	incontáveis	2925	1	irma
2886	1	incríveis	2926	1	irmaões
2887	1	incrível	2927	1	irritando
2888	1	indereso	2928	1	irá
2889	1	indignação	2929	1	iríam
2890	1	indiretas	2930	1	isaacs
2891	1	inesperado	2931	1	isgoto
2892	1	inesquecíveis	2932	1	iso
2893	1	inferior	2933	1	istef
2894	1	infermeiros	2934	1	istão
2895	1	inferno	2935	1	janela
2896	1	influenciando	2936	1	janelas
2897	1	inforcou	2937	1	januza
2898	1	injeçado	2938	1	japão
2899	1	inprego	2939	1	jardim
2900	1	insegurança	2940	1	jason
2901	1	insina	2941	1	jenifer
2902	1	inspirado	2942	1	joelho
2903	1	inspirou	2943	1	jogadores
2904	1	insuportáveis	2944	1	jogam
2905	1	inteligência	2945	1	jogaçem
2906	1	intensa	2946	1	jogem
2907	1	intenção	2947	1	jogos
2908	1	intera	2948	1	jogão
2909	1	intercambio	2949	1	joia
2910	1	interessante	2950	1	jolie

2951	1	jornalistas	2991	1	lawernnce
2952	1	jovem	2992	1	leais
2953	1	judgam	2993	1	legals
2954	1	judgamentos	2994	1	leio
2955	1	judgamos	2995	1	leis
2956	1	judgaram	2996	1	lembra
2957	1	judgasem	2997	1	lembrados
2958	1	judguem	2998	1	lembrando
2959	1	julia	2999	1	lembranças
2960	1	juliana	3000	1	lembrar
2961	1	july	3001	1	lembrei
2962	1	juntamente	3002	1	lendário
2963	1	juntas	3003	1	lenhadores
2964	1	justas	3004	1	leonardo
2965	1	justiça	3005	1	levada
2966	1	jutsu	3006	1	levaram
2967	1	juventus	3007	1	levarem
2968	1	jão	3008	1	levaria
2969	1	júlia	3009	1	levariam
2970	1	kkkk	3010	1	levava
2971	1	kão	3011	1	leveí
2972	1	labrador	3012	1	lgbt
2973	1	ladrões	3013	1	lhe
2974	1	lagoas	3014	1	liberados
2975	1	lagos	3015	1	liberdade
2976	1	lampâdas	3016	1	liberdades
2977	1	lancherias	3017	1	libertadores
2978	1	lanches	3018	1	ligar
2979	1	lanchonete	3019	1	ligaram
2980	1	lanjonete	3020	1	ligava
2981	1	lançarem	3021	1	lily
2982	1	larga	3022	1	limpando
2983	1	largar	3023	1	limpas
2984	1	largarem	3024	1	limpassem
2985	1	largou	3025	1	limpava
2986	1	larissa	3026	1	limpeza
2987	1	las	3027	1	limpos
2988	1	lasaro	3028	1	linda
2989	1	lasso	3029	1	lindos
2990	1	latinha	3030	1	linpar

3031	1	linpe	3071	1	mandam
3032	1	lipas	3072	1	mandar
3033	1	lipos	3073	1	mandaram
3034	1	literalmente	3074	1	mandava
3035	1	livrar	3075	1	mandou
3036	1	livremente	3076	1	maneiras
3037	1	livro	3077	1	manhãs
3038	1	lição	3078	1	manifestações
3039	1	locas	3079	1	manifestos
3040	1	lock	3080	1	mantemos
3041	1	lojinha	3081	1	manter
3042	1	longa	3082	1	mantido
3043	1	longe	3083	1	maravilhosas
3044	1	longos	3084	1	marcadas
3045	1	lotadas	3085	1	marcando
3046	1	lua	3086	1	marcar
3047	1	luis	3087	1	marcou
3048	1	lula	3088	1	maria
3049	1	luxuosas	3089	1	marido
3050	1	luz	3090	1	marrem
3051	1	lésbicas	3091	1	massacre
3052	1	língua	3092	1	matam
3053	1	machucar	3093	1	matamento
3054	1	machucou	3094	1	matamos
3055	1	machuquei	3095	1	mataria
3056	1	maconha	3096	1	matação
3057	1	maduros	3097	1	matem
3058	1	magia	3098	1	matematica
3059	1	magica	3099	1	matemática
3060	1	magina	3100	1	materia
3061	1	magoada	3101	1	matou
3062	1	maiores	3102	1	matá
3063	1	malik	3103	1	matão
3064	1	malstratos	3104	1	maustratar
3065	1	maltratado	3105	1	maximo
3066	1	mama	3106	1	mccrory
3067	1	mamãe	3107	1	mecanico
3068	1	man	3108	1	medalha
3069	1	manda	3109	1	medalhas
3070	1	mandada	3110	1	mederam

3111	1	medicamentos	3151	1	minquen
3112	1	medicina	3152	1	minutos
3113	1	megan	3153	1	minima
3114	1	meia	3154	1	mio
3115	1	melhora	3155	1	mique
3116	1	melhoras	3156	1	miseria
3117	1	melhorzinhas	3157	1	missisipi
3118	1	melor	3158	1	misteriosa
3119	1	membros	3159	1	misturada
3120	1	memoria	3160	1	misturando
3121	1	mendingo	3161	1	misturaria
3122	1	meneses	3162	1	moda
3123	1	menhos	3163	1	modificar
3124	1	menor	3164	1	moedas
3125	1	mentais	3165	1	monte
3126	1	mente	3166	1	monumentos
3127	1	mentes	3167	1	moonte
3128	1	mentigo	3168	1	moradia
3129	1	mentiras	3169	1	moram
3130	1	mercadinho	3170	1	moravamos
3131	1	mercadorias	3171	1	morem
3132	1	merdas	3172	1	mores
3133	1	merecem	3173	1	morrar
3134	1	mereceu	3174	1	morrem
3135	1	mes	3175	1	morrer
3136	1	mesa	3176	1	morreram
3137	1	mesas	3177	1	morreria
3138	1	meteu	3178	1	morresseem
3139	1	meticos	3179	1	morreu
3140	1	metidos	3180	1	mortais
3141	1	micel	3181	1	mortalidade
3142	1	mickol	3182	1	morto
3143	1	micy	3183	1	mos
3144	1	mikey	3184	1	mostraria
3145	1	milenio	3185	1	mostrava
3146	1	mimada	3186	1	mostrou
3147	1	mimadas	3187	1	motinho
3148	1	min	3188	1	motivador
3149	1	minimas	3189	1	motivadoras
3150	1	minimo	3190	1	motivações

3191	1	moto	3231	1	necessidades
3192	1	mudala	3232	1	necessita
3193	1	mudanças	3233	1	necessitada
3194	1	mudassem	3234	1	necessitados
3195	1	mudava	3235	1	necessitam
3196	1	mudo	3236	1	necessitarem
3197	1	muítaaa	3237	1	negocios
3198	1	multa	3238	1	negros
3199	1	multidão	3239	1	nerd
3200	1	mundial	3240	1	nervosa
3201	1	munta	3241	1	nervosas
3202	1	muntas	3242	1	nervosismo
3203	1	muto	3243	1	nese
3204	1	muíto	3244	1	nessas
3205	1	má	3245	1	nessecitados
3206	1	mágoas	3246	1	netiflix
3207	1	mães	3247	1	nickolas
3208	1	média	3248	1	ningue
3209	1	médicamentos	3249	1	noah
3210	1	médicos	3250	1	nois
3211	1	méxico	3251	1	nomeado
3212	1	mêses	3252	1	nomes
3213	1	mínimo	3253	1	nordeste
3214	1	nacionalidade	3254	1	norte
3215	1	nadas	3255	1	nosssa
3216	1	namorar	3256	1	nosça
3217	1	naquele	3257	1	notem
3218	1	nariz	3258	1	noticia
3219	1	naruto	3259	1	noticias
3220	1	nascem	3260	1	notinha
3221	1	nascidos	3261	1	notícias
3222	1	natalia	3262	1	novela
3223	1	nativa	3263	1	noção
3224	1	naverda	3264	1	nu
3225	1	navil	3265	1	nuca
3226	1	nação	3266	1	numca
3227	1	nct	3267	1	número
3228	1	ne	3268	1	números
3229	1	necessaria	3269	1	obem
3230	1	necessario	3270	1	obotunidade

3271	1	obra	3311	1	organização
3272	1	obrigação	3312	1	orgulhe
3273	1	obrigada	3313	1	orgulho
3274	1	obrigado	3314	1	orgânicos
3275	1	obrigados	3315	1	oscar
3276	1	obrigar	3316	1	osso
3277	1	obrigação	3317	1	ótima
3278	1	observam	3318	1	otimismo
3279	1	observaram	3319	1	ótimo
3280	1	observações	3320	1	ouve
3281	1	obviamente	3321	1	ouvi
3282	1	oceanos	3322	1	oxigênio
3283	1	oculta	3323	1	p
3284	1	ocupadas	3324	1	padaria
3285	1	ofender	3325	1	padarias
3286	1	oferecer	3326	1	padrão
3287	1	oher	3327	1	padrões
3288	1	oi	3328	1	paeva
3289	1	oitavas	3329	1	paga
3290	1	oito	3330	1	pagam
3291	1	olhar	3331	1	pagamentos
3292	1	olharam	3332	1	pagando
3293	1	olharmos	3333	1	pagaram
3294	1	olheiros	3334	1	pagariam
3295	1	olhou	3335	1	pagará
3296	1	omicídios	3336	1	paisagens
3297	1	onesta	3337	1	palavram
3298	1	onestos	3338	1	palavras
3299	1	ongs	3339	1	palavrinha
3300	1	onéstio	3340	1	palhaçada
3301	1	operação	3341	1	palmares
3302	1	opinar	3342	1	pandido
3303	1	oponente	3343	1	papa
3304	1	oponentes	3344	1	papelão
3305	1	oportunidad	3345	1	papo
3306	1	oportunidades	3346	1	par
3307	1	opresidente	3347	1	parada
3308	1	opção	3348	1	parados
3309	1	opínião	3349	1	paraná
3310	1	orfanato	3350	1	pararem

3351	1	paras	3391	1	pegado
3352	1	parasse	3392	1	pegalos
3353	1	parceiro	3393	1	pegamos
3354	1	parceiros	3394	1	pegaria
3355	1	parcelado	3395	1	pegh
3356	1	parecem	3396	1	pego
3357	1	parecer	3397	1	peito
3358	1	parecia	3398	1	pele
3359	1	parede	3399	1	pensado
3360	1	pareseu	3400	1	pensamentos
3361	1	pareça	3401	1	pensativa
3362	1	parga	3402	1	pensei
3363	1	paro	3403	1	pensãm
3364	1	partamento	3404	1	pequenos
3365	1	participar	3405	1	perca
3366	1	participaram	3406	1	percebam
3367	1	particularmente	3407	1	perceberam
3368	1	partir	3408	1	perde
3369	1	partiria	3409	1	perdem
3370	1	partiu	3410	1	perdemos
3371	1	parto	3411	1	perderia
3372	1	pasar	3412	1	perdeu
3373	1	passamos	3413	1	perfeita
3374	1	passaria	3414	1	perguntam
3375	1	passava	3415	1	perguntou
3376	1	passavamos	3416	1	periferia
3377	1	passe	3417	1	perigosas
3378	1	passei	3418	1	perigoso
3379	1	passeios	3419	1	perito
3380	1	passer	3420	1	periodos
3381	1	passo	3421	1	permitido
3382	1	patos	3422	1	perna
3383	1	patoscidym	3423	1	perseguido
3384	1	peculiaridades	3424	1	perseguidos
3385	1	pedindo	3425	1	persistir
3386	1	pedofilia	3426	1	perssebe
3387	1	pedra	3427	1	pertencer
3388	1	pedrinhas	3428	1	pesados
3389	1	pedrinho	3429	1	pesoas
3390	1	pedro	3430	1	piadinhas



3431	1	piano	3471	1	ponta
3432	1	picape	3472	1	popreses
3433	1	pichados	3473	1	porem
3434	1	pichar	3474	1	porquê
3435	1	pietra	3475	1	porra
3436	1	pintar	3476	1	porrada
3437	1	pinxar	3477	1	portao
3438	1	piores	3478	1	portas
3439	1	piqueno	3479	1	português
3440	1	piratas	3480	1	portão
3441	1	pirâmide	3481	1	porém
3442	1	pisei	3482	1	poso
3443	1	pistas	3483	1	possamos
3444	1	pixadas	3484	1	possiveis
3445	1	planetas	3485	1	possuir
3446	1	planos	3486	1	possuu
3447	1	plantar	3487	1	posta
3448	1	plantamos	3488	1	postar
3449	1	plastico	3489	1	posti
3450	1	platico	3490	1	posto
3451	1	playstation	3491	1	posição
3452	1	po	3492	1	potam
3453	1	poche	3493	1	potem
3454	1	podem	3494	1	potencia
3455	1	podera	3495	1	poucos
3456	1	poderam	3496	1	poupar
3457	1	poderão	3497	1	poupava
3458	1	podessemos	3498	1	povoado
3459	1	podiam	3499	1	pracinha
3460	1	podres	3500	1	prancha
3461	1	politica	3501	1	prassas
3462	1	politico	3502	1	prassinhas
3463	1	poluido	3503	1	praticamente
3464	1	poluidos	3504	1	praticar
3465	1	poluindo	3505	1	praticula
3466	1	poluído	3506	1	praças
3467	1	poluídos	3507	1	preciosa
3468	1	poluísão	3508	1	precisar
3469	1	políciamento	3509	1	precisarmos
3470	1	ponha	3510	1	precisasse

3511	1	precisava	3551	1	prisões
3512	1	precisavamos	3552	1	privados
3513	1	precisem	3553	1	probres
3514	1	preciso	3554	1	produtos
3515	1	preconceitooso	3555	1	profesores
3516	1	preferem	3556	1	profissional
3517	1	preferido	3557	1	profundas
3518	1	prefeto	3558	1	progetos
3519	1	prefiro	3559	1	progresso
3520	1	pregeudicar	3560	1	proibida
3521	1	preguiçosas	3561	1	promete
3522	1	prejudicados	3562	1	pronta
3523	1	prejudique	3563	1	pronto
3524	1	premiação	3564	1	prontos
3525	1	prendam	3565	1	proposito
3526	1	prendem	3566	1	proposta
3527	1	prendão	3567	1	propria
3528	1	preocupa	3568	1	proteger
3529	1	preocupasem	3569	1	proternizado
3530	1	preocupo	3570	1	protestos
3531	1	presas	3571	1	provas
3532	1	prescisam	3572	1	provavel
3533	1	presentes	3573	1	proxímo
3534	1	presidiário	3574	1	prá
3535	1	preso	3575	1	prédio
3536	1	pressisamos	3576	1	próprios
3537	1	preta	3577	1	próxima
3538	1	pretendo	3578	1	psicologia
3539	1	preucupo	3579	1	pude
3540	1	preucurando	3580	1	pudece
3541	1	previsão	3581	1	pudecem
3542	1	preça	3582	1	pudese
3543	1	primas	3583	1	pudessemos
3544	1	primeiramente	3584	1	pulamos
3545	1	primeiras	3585	1	pulando
3546	1	primero	3586	1	pulava
3547	1	princípal	3587	1	pulicias
3548	1	princípalmente	3588	1	puliticos
3549	1	priquedos	3589	1	pulmão
3550	1	prisioneiro	3590	1	puxar

3591	1	puxem	3631	1	reais
3592	1	pátio	3632	1	reaparecendo
3593	1	pão	3633	1	reação
3594	1	péssimo	3634	1	rebeldes
3595	1	pública	3635	1	recado
3596	1	público	3636	1	recaídas
3597	1	quadro	3637	1	recebam
3598	1	quak	3638	1	recebemos
3599	1	qualificado	3639	1	receber
3600	1	quaquer	3640	1	receberam
3601	1	quardo	3641	1	receberem
3602	1	quarta	3642	1	receberia
3603	1	quartel	3643	1	recebermos
3604	1	quebrados	3644	1	recebi
3605	1	quebram	3645	1	reciclar
3606	1	quebrei	3646	1	recipiente
3607	1	quebrou	3647	1	recomemeçar
3608	1	queimada	3648	1	recomendaria
3609	1	queiram	3649	1	reconhecer
3610	1	queiroz	3650	1	reconpença
3611	1	queis	3651	1	recorde
3612	1	quela	3652	1	recramando
3613	1	queimar	3653	1	recuperar
3614	1	queremos	3654	1	recupere
3615	1	queridas	3655	1	recusar
3616	1	quietas	3656	1	redentor
3617	1	quimadas	3657	1	redes
3618	1	quince	3658	1	redor
3619	1	quisecem	3659	1	reduzem
3620	1	quiserem	3660	1	reduzir
3621	1	quão	3661	1	reencontra
3622	1	r	3662	1	reencontrando
3623	1	racismos	3663	1	reencontraram
3624	1	racista	3664	1	refeição
3625	1	ralph	3665	1	referencias
3626	1	ramos	3666	1	referente
3627	1	rancor	3667	1	refletirem
3628	1	rapoza	3668	1	reformular
3629	1	raquetes	3669	1	reformas
3630	1	rasga	3670	1	refém

3671	1	refêns	3711	1	restaurantes
3672	1	regeitado	3712	1	resumir
3673	1	região	3713	1	retardados
3674	1	reino	3714	1	retiraria
3675	1	rejeitado	3715	1	retomar
3676	1	rejeitados	3716	1	retornando
3677	1	relação	3717	1	retornar
3678	1	relevar	3718	1	reunen
3679	1	remedíos	3719	1	reuniram
3680	1	remunerária	3720	1	reunisse
3681	1	renan	3721	1	rica
3682	1	rendendo	3722	1	rickman
3683	1	renderam	3723	1	ridículo
3684	1	rendibol	3724	1	rindo
3685	1	reparos	3725	1	riquezas
3686	1	repeitar	3726	1	riria
3687	1	repetir	3727	1	risada
3688	1	replantar	3728	1	risonhos
3689	1	replantio	3729	1	roba
3690	1	reportagem	3730	1	robando
3691	1	reserva	3731	1	robin
3692	1	resgatando	3732	1	roda
3693	1	resgate	3733	1	rodei
3694	1	resinerar	3734	1	rodeiam
3695	1	resistir	3735	1	rodinha
3696	1	resolve	3736	1	rodou
3697	1	resolver	3737	1	rosa
3698	1	resolveu	3738	1	rosto
3699	1	resolvido	3739	1	rota
3700	1	respeinta	3740	1	roubarem
3701	1	respeitacem	3741	1	roube
3702	1	respeitado	3742	1	roubou
3703	1	respeitam	3743	1	roubãom
3704	1	respeitasem	3744	1	rumo
3705	1	respeitasse	3745	1	rural
3706	1	respiração	3746	1	s
3707	1	respondendo	3747	1	sabado
3708	1	responder	3748	1	sabiamos
3709	1	respondeu	3749	1	saci
3710	1	responsabilidade	3750	1	saem

3751	1	saiam	3791	1	semi
3752	1	saiamos	3792	1	semprena
3753	1	saido	3793	1	semten
3754	1	saiem	3794	1	sen
3755	1	sairão	3795	1	senas
3756	1	saião	3796	1	sensação
3757	1	saldaveis	3797	1	sentia
3758	1	salto	3798	1	sentimos
3759	1	salvam	3799	1	separadas
3760	1	salvarem	3800	1	separado
3761	1	salvão	3801	1	separados
3762	1	salários	3802	1	separamos
3763	1	samana	3803	1	separando
3764	1	saniamento	3804	1	separassem
3765	1	santo	3805	1	separava
3766	1	sao	3806	1	sequestrado
3767	1	sarandi	3807	1	sequestram
3768	1	saudades	3808	1	sequestrar
3769	1	saudavel	3809	1	sequestrão
3770	1	saírem	3810	1	serio
3771	1	saúdes	3811	1	serião
3772	1	scorpious	3812	1	sertesa
3773	1	seco	3813	1	serto
3774	1	secos	3814	1	serve
3775	1	secreto	3815	1	serviam
3776	1	secuestre	3816	1	serviços
3777	1	sede	3817	1	será
3778	1	segredo	3818	1	seríamos
3779	1	seguem	3819	1	setarem
3780	1	sequestra	3820	1	sexo
3781	1	sequestrado	3821	1	sexta
3782	1	segundos	3822	1	shopping
3783	1	segunransa	3823	1	shoppings
3784	1	seguras	3824	1	show
3785	1	seguros	3825	1	sidadão
3786	1	segurou	3826	1	sinal
3787	1	seis	3827	1	sinta
3788	1	sejamos	3828	1	sintir
3789	1	seje	3829	1	sloty
3790	1	sejão	3830	1	smith

3831	1	sniper	3871	1	sumiu
3832	1	sobra	3872	1	super
3833	1	sobrecarreguem	3873	1	supera
3834	1	sobreu	3874	1	supermercado
3835	1	sobrevivencia	3875	1	supor
3836	1	sociais	3876	1	suspense
3837	1	social	3877	1	susto
3838	1	sofrem	3878	1	suzano
3839	1	sofria	3879	1	sábados
3840	1	sofrida	3880	1	sáude
3841	1	sogar	3881	1	série
3842	1	soldados	3882	1	sério
3843	1	solidarios	3883	1	sótam
3844	1	solo	3884	1	tachau
3845	1	soltalos	3885	1	tacos
3846	1	sonhadores	3886	1	tals
3847	1	sonham	3887	1	tamto
3848	1	sonharem	3888	1	tanben
3849	1	sonolentos	3889	1	tao
3850	1	sonoramente	3890	1	tarefa
3851	1	soquero	3891	1	tchau
3852	1	sorrisos	3892	1	tecidos
3853	1	sota	3893	1	tecnicas
3854	1	soube	3894	1	tecnolôgia
3855	1	sozinha	3895	1	tejada
3856	1	sozinho	3896	1	tema
3857	1	sra	3897	1	temas
3858	1	strupo	3898	1	tempestade
3859	1	ststentava	3899	1	tempos
3860	1	sub	3900	1	ten
3861	1	subi	3901	1	tenhão
3862	1	substituido	3902	1	tenis
3863	1	substituta	3903	1	tentassem
3864	1	subterrâneo	3904	1	terceiro
3865	1	sucedido	3905	1	tereno
3866	1	sucesso	3906	1	termina
3867	1	suicidios	3907	1	terminou
3868	1	sujas	3908	1	termo
3869	1	sujos	3909	1	termos
3870	1	sumir	3910	1	terras

3911	1	terá	3951	1	trabalhador
3912	1	terão	3952	1	trabalhando
3913	1	teríamos	3953	1	trabalhava
3914	1	teste	3954	1	trafego
3915	1	teto	3955	1	tragédia
3916	1	tevolva	3956	1	tragédias
3917	1	thais	3957	1	trancado
3918	1	thiaguinho	3958	1	tranquilamente
3919	1	ti	3959	1	tranquilo
3920	1	tia	3960	1	transformar
3921	1	tijolino	3961	1	transformaria
3922	1	tio	3962	1	transformá
3923	1	tiraria	3963	1	transporte
3924	1	tiretoris	3964	1	trata
3925	1	tirotei	3965	1	tratadas
3926	1	tiroteio	3966	1	tratamentos
3927	1	tiroteiros	3967	1	traz
3928	1	tirou	3968	1	traçam
3929	1	titularidade	3969	1	treinado
3930	1	tivemos	3970	1	treinalo
3931	1	tivese	3971	1	treinamento
3932	1	tivesem	3972	1	treinando
3933	1	tivesen	3973	1	tres
3934	1	tivisse	3974	1	trilha
3935	1	tocar	3975	1	trimestre
3936	1	tom	3976	1	tristemente
3937	1	toma	3977	1	tristezas
3938	1	tomado	3978	1	trolls
3939	1	tomando	3979	1	trono
3940	1	tomei	3980	1	troque
3941	1	tomem	3981	1	troquem
3942	1	tomemos	3982	1	tua
3943	1	tomo	3983	1	tue
3944	1	tomou	3984	1	tunem
3945	1	top	3985	1	turne
3946	1	tornaram	3986	1	twitter
3947	1	tornarão	3987	1	tâm
3948	1	tornasse	3988	1	têm
3949	1	torne	3989	1	tô
3950	1	total	3990	1	túnel

3991	1	umbigo	4031	1	venha
3992	1	umilde	4032	1	verba
3993	1	umildes	4033	1	verbal
3994	1	un	4034	1	verdadeiro
3995	1	una	4035	1	vereadores
3996	1	unam	4036	1	verem
3997	1	unica	4037	1	vestibulares
3998	1	unir	4038	1	vestir
3999	1	upas	4039	1	viagem
4000	1	urinarem	4040	1	viagens
4001	1	usa	4041	1	viajou
4002	1	usinas	4042	1	vicios
4003	1	uso	4043	1	videogame
4004	1	utilizou	4044	1	videos
4005	1	utima	4045	1	vidro
4006	1	vaise	4046	1	viemos
4007	1	valendo	4047	1	vieram
4008	1	valiosa	4048	1	vigilante
4009	1	valiosas	4049	1	vingança
4010	1	valiozas	4050	1	vingar
4011	1	valkiria	4051	1	vio
4012	1	valores	4052	1	virando
4013	1	valorizar	4053	1	virei
4014	1	valões	4054	1	visitas
4015	1	vaso	4055	1	visse
4016	1	veem	4056	1	vissos
4017	1	vegano	4057	1	vistos
4018	1	vegetariano	4058	1	vitória
4019	1	veio	4059	1	vivam
4020	1	veis	4060	1	vive
4021	1	veja	4061	1	vivenciaram
4022	1	velha	4062	1	viveram
4023	1	velhinho	4063	1	viverem
4024	1	velhos	4064	1	viviam
4025	1	vencer	4065	1	vizinhança
4026	1	venceríamos	4066	1	vizinho
4027	1	venceu	4067	1	voar
4028	1	venci	4068	1	vocalistas
4029	1	vendura	4069	1	voce
4030	1	veneno	4070	1	voleibol



4071	1	voltam	4111	1	ódio
4072	1	voltando	4112	1	ônibus
4073	1	voltarei	4113	1	último
4074	1	voltasse			
4075	1	voltava			
4076	1	voltou			
4077	1	vomito			
4078	1	vorom			
4079	1	votar			
4080	1	voutan			
4081	1	voutaram			
4082	1	voz			
4083	1	vulgo			
4084	1	vãrn			
4085	1	washington			
4086	1	werneck			
4087	1	wili			
4088	1	will			
4089	1	william			
4090	1	willians			
4091	1	x			
4092	1	xingamento			
4093	1	xingar			
4094	1	xingão			
4095	1	xis			
4096	1	xuxa			
4097	1	yans			
4098	1	yasmim			
4099	1	youtube			
4100	1	zayn			
4101	1	zoam			
4102	1	às			
4103	1	ámbulancias			
4104	1	ápos			
4105	1	área			
4106	1	çadroses			
4107	1	época			
4108	1	ética			
4109	1	íamos			
4110	1	óbvio			