


unesp  UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Faculdade de Ciências e Letras
Campus de Araraquara - SP

DANIELE TREVELIN DONATO-LUZ

**LEVANTAMENTO DO LÉXICO MAIS
FREQUENTE EM ARTIGOS JORNALÍSTICOS
DA ÁREA DE EDUCAÇÃO: uso de *corpora* para
indicação de correspondência português-inglês**



ARARAQUARA – S.P.
2019

DANIELE TREVELIN DONATO-LUZ

**LEVANTAMENTO DO LÉXICO MAIS FREQUENTE
EM ARTIGOS JORNALÍSTICOS DA ÁREA DE
EDUCAÇÃO: uso de *corpora* para indicação de
correspondência português-inglês**

Dissertação de Mestrado apresentada ao Programa de Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – Unesp/Araraquara, como requisito para obtenção do título de Mestre em Linguística e Língua Portuguesa.

Linha de pesquisa: Estudos do léxico

Orientador: Prof. Dr. Celso Fernando Rocha

ARARAQUARA – S.P.
2019

Donato-Luz, Daniele Trevelin
Levantamento do léxico mais frequente em artigos
jornalísticos da área de educação: uso de corpora para
indicação de correspondência português-inglês /
Daniele Trevelin Donato-Luz – 2019
108 f.

Dissertação (Mestrado em Linguística e Língua
Portuguesa) – Universidade Estadual Paulista "Júlio
de Mesquita Filho", Faculdade de Ciências e Letras
(Campus Araraquara)

Orientador: Celso Fernando Rocha

1. Linguística de Corpus. 2. Artigos jornalísticos. 3.
Educação. 4. Glossário bilíngue. 5. Equivalência
português-inglês. I. Título.

DANIELE TREVELIN DONATO-LUZ

**LEVANTAMENTO DO LÉXICO MAIS FREQUENTE
EM ARTIGOS JORNALÍSTICOS DA ÁREA DE
EDUCAÇÃO: uso de *corpora* para indicação de
correspondência português-inglês**

Dissertação de Mestrado apresentada ao Programa de Linguística e Língua Portuguesa da Faculdade de Ciências e Letras – Unesp/Araraquara, como requisito para obtenção do título de Mestre em Linguística e Língua Portuguesa.

Linha de pesquisa: Estudos do léxico

Orientador: Prof. Dr. Celso Fernando Rocha

Data da defesa: 29/04/2019

MEMBROS COMPONENTES DA BANCA EXAMINADORA:

Presidente e Orientador: Prof. Dr. Celso Fernando Rocha
Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP/Araraquara

Membro Titular: Prof. Dra. Camila Höfling
Universidade Federal de São Carlos - UFSCar

Membro Titular: Prof. Dra. Regiani Aparecida Santos Zacarias
Universidade Estadual Paulista “Júlio Mesquita Filho” – UNESP/Assis

Local: Universidade Estadual Paulista
Faculdade de Ciências e Letras
UNESP – Campus de Araraquara

Aos meus pais, Célia e Saturnino, meu refúgio e razão do meu viver.
Ao meu marido Heric, razão da minha felicidade.
A minha irmã Dayane, minha eterna companheira.

AGRADECIMENTOS

À Deus e Nossa Senhora Aparecida por me sustentar, guiar e proteger de todo mal e perigo. Por terem me acompanhado em cada viagem e em cada passo para a realização dessa pesquisa.

Aos meus pais, Maria Célia e José Saturnino pelo amor incondicional, pelo exemplo, pela dedicação e por não medirem esforços para que eu chegasse até aqui. Devo tudo a vocês e que eu possa ser um pouco do que vocês foram e são para mim. Minha Maria e meu José, exemplo de família, de bondade, de coragem.

Ao meu marido, Heric Willian, que sempre acreditou em mim, mesmo quando nem eu acreditava. Sou grata pelo seu incentivo, por compreender minha ausência e a série de momentos em que o nervosismo se fazia presente. Você é minha calma durante toda turbulência. Também o agradeço pela ajuda com as tabelas no Excel, sem você eu não teria conseguido organizar os dados.

À minha irmã Dayane, minha caçulinha, a mais brava do mundo, mas a quem eu amo com todas as forças. Não entende porque vivo entre os livros, mas me apoia em qualquer decisão.

Aos meus sogros, Nicilene e Ademir, por compreenderem as várias vezes em que não pude estar presente, por me apoiarem e me adotarem como parte da família.

À minha cunhada Thais, a irmã mais velha, pela companhia e compreensão durante essa caminhada

Ao professor e orientador Dr. Celso Fernando Rocha pela paciência, sabedoria e conhecimento transmitidos no decorrer dessa árdua caminhada. Sem seus ensinamentos, esse trabalho não teria sido possível.

À minha amiga Mirian, parceira em todos os momentos desse percurso, dividindo comigo não só sua amizade, como seu tempo, sua casa, seus ideais. Passamos por muitos obstáculos, mas nunca desistimos. Que nossa amizade possa continuar eternamente.

À minha amiga Aline, por não ter desistido de mim, após as inúmeras vezes em que recusei seus convites para pedalar. E agradeço por me ouvir e aconselhar quando tudo parecia perdido.

À professora Regiani, carinhosamente chamada de Reça, por ter plantado a sementinha da pesquisa durante a graduação, me apresentando esse universo e guiado meus primeiros passos na pesquisa acadêmica. Seus ensinamentos foram cruciais para a formação da pessoa que sou hoje.

Às professoras Regiani e Adriane pelas contribuições durante a banca de Qualificação, essenciais para o aprimoramento da pesquisa.

Aos membros suplentes da banca de defesa, Prof. Dra. Nildiceia Aparecida Rocha e Prof. Dra. Paula Tavares Pinto, por terem me desejado sucesso nessa etapa.

Aos membros da banca de defesa, professoras Camila Höfling e Regiani Zacarias pela leitura minuciosa, comentários e sugestões durante a banca de defesa. Agradeço imensamente o cuidado e tempo que dedicaram ao texto.

RESUMO

A educação é tema comum e presente em toda sociedade, relaciona-se a diversos momentos da vida cotidiana individual e coletiva e está presente em muitas das informações divulgadas diariamente em textos jornalísticos e na mídia em geral. Dessa forma, aliados ao arcabouço teórico-metodológico da Linguística de *Corpus* (BERBER SARDINHA, 2000, 2004; TOGNINI-BONELLI, 2001, BIBER, 1998, SINCLAIR, 1991, 1996; BEVILACQUA, 2013), a alguns princípios da Terminologia e Terminografia (AUBERT, 1998; BARROS, 2004; KRIEGER & FINATTO, 2004, BARBOSA, 1999) e pressupostos da Lexicologia e Lexicografia (BIDERMAN, 1987, 1996, 1998, 2001; WELKER, 2004) temos como objetivos gerais: 1) identificar os termos com maior chavidade na área de educação em língua portuguesa e língua inglesa a partir de *corpora* de artigos jornalísticos retirados de quatro jornais, sendo dois brasileiros (*O Globo* e *O Estadão*) e dois norte-americanos (*The New York Times* e *Los Angeles Times*) e 2) Estabelecer a relação de correspondência léxico-semântica português-inglês dos termos estatisticamente significativos. Na primeira etapa da pesquisa, identificamos as palavras mais frequentes para compor uma lista de termos em língua portuguesa (LP) e língua inglesa (LI). Para tanto, utilizamos quatro *corpora* comparáveis, compilados a partir dos jornais *O Estadão*, *O Globo*, *The New York Times* e *Los Angeles Times*. Cada corpus é composto de cem artigos publicados no período de junho a agosto de 2017, nas versões online dos jornais. Por meio das ferramentas da Linguística de *Corpus*, foram selecionados 41 termos simples, da classe dos substantivos comuns em língua portuguesa. Na segunda etapa, buscamos analisar contrastivamente a relação de correspondência semântica interlínguas português-inglês dos termos, observando as questões culturais e relacionadas aos sistemas educacionais brasileiro e americano. A partir dos termos simples, verificamos também composições sintagmáticas (BARROS, 2004) e suas correspondências. Os resultados obtidos nas análises motivaram a proposta de glossário bilíngue a ser hospedado em ambiente virtual e poderão servir para a compilação de materiais terminológicos na área de Educação, tendo em vista a escassez de obras neste setor, especialmente de obras bilíngues. Os mesmos resultados podem contribuir para estudantes de língua estrangeira, tradutores e interessados.

Palavras-chave: Linguística de *Corpus*. Artigos jornalísticos. Educação. Glossário Bilíngue. Equivalência português-inglês.

ABSTRACT

Education is a usual topic and it is present in every society. It is related to several daily routine's moments, in the individual or collective life, and it is in the information published daily in news articles and in the media in general. So, together with the theoretical-methodological framework of Corpus Linguistics (BERBER SARDINHA, 2000, 2004, TOGNINI-BONELLI, 2001, BIBER, 1998, SINCLAIR, BEVILACQUA, 2013), some principles of Terminology and Terminography (AUBERT, 1998; BARROS, 2004; KRIEGER & FINATTO, 2004, BARBOSA, 1999) and assumptions of Lexicology and Lexicography (BIDERMAN, 1987, 1996, 1998, 2001; WELKER, 2004), we aim to: 1) to find out the most frequent words from education taken from news articles, written in Portuguese and English. The articles are from two Brazilian newspapers (*O Globo* and *O Estadão*) and two north-American newspapers (*The New York Times* and *Los Angeles Times*); 2) to establish the Portuguese-English lexical-semantic correspondence of the terms statistically meaningful. In the first step, we identify the most frequent words to make a list with Portuguese terms and English terms. We used four comparable corpora from *O Estadão*, *O Globo*, *The New York Times* and *Los Angeles Times*. Each corpus has a hundred articles published from June to August of 2017. Using the Corpus Linguistics tools, we selected 41 Portuguese nouns. On the second step, we aim to analyze the interlanguage correspondence; we studied cultural issues related to the educational systems of both countries. We also found syntagmatic compositions (BARROS, 2004) and their correspondences. The data may serve to compile terminological materials in the area of education, due to the lack of works in this subject, especially bilingual. The information may contribute to foreign language students, translators or public in general.

Keywords: Corpus Linguistics. News articles. Education. Bilingual Glossary. Portuguese-English equivalence.

LISTA DE TABELAS

Tabela 1 - Palavras-chave substantivas	61
Tabela 2 - Quantidade de substantivos selecionados	61

LISTA DE QUADROS

Quadro 1 - Estrutura do sistema educacional brasileiro	24
Quadro 2 - Estrutura do sistema educacional americano	26
Quadro 3 - Critérios para compilação dos corpora	36
Quadro 4 - Sistematização das áreas das Ciências do Léxico	53
Quadro 5 – Função das ferramentas do WordSmith Tools e propósito na pesquisa	57
Quadro 6 - Termos selecionados para busca de correspondentes	62
Quadro 7 - Substantivo em português e correspondente em inglês	64
Quadro 8 - Procedimentos adotados durante a análise	66
Quadro 9 - Termos não encontrados no corpus de língua inglesa	78

LISTA DE FIGURAS

Figura 1 - Verbetes 'educação' no dicionário Aurélio	13
Figura 2 - Verbetes 'education' no dicionário Longman Dictionary of Contemporary English	20
Figura 3 - Tela inicial do WordSmith Tools, versão 6.0	39
Figura 4 - Lista de palavras dos jornais O Globo e O Estadão	40
Figura 5 - Listas de palavras dos jornais Los Angeles Times e New York Times	41
Figura 6 - Linhas de concordância a partir da busca por 'ensino'	44
Figura 7 - Verbetes 'bolsa' no Dicionário Aurélio com destaque às siglas de especialidade	51
Figura 8 - Siglas no Dicionário Aurélio e destaque para a sigla de 'educação'	52
Figura 9 - Seleção dos textos para criação da WordList	58
Figura 10 - Lista de palavras do jornal O Estadão organizadas por ordem de frequência	59
Figura 11 - Lista de palavras-chave do jornal O Estadão	60
Figura 12 - Busca pela palavra 'ensino' na ferramenta Concord	63
Figura 13 - Reorganização das composições sintagmáticas	71
Figura 14 - Comparação dos sistemas educacionais	72
Figura 15 - PrintScreen do site Glossário Bilíngue de Educação	81
Figura 16 - PrintScreen da apresentação do termo <i>educação</i> no Glossário Bilíngue de Educação	82

LISTA DE ABREVIATURAS E SIGLAS

LC	Linguística de Corpus
TCT	Teoria Comunicativa da Terminologia
TGT	Teoria Geral da Terminologia
NEC	Não Encontrado no Corpus

SUMÁRIO

INTRODUÇÃO	16
1. A EDUCAÇÃO COMO TEMA PARA ESTUDO DA CORRESPONDÊNCIA LÉXICO-SEMÂNTICA PORTUGUÊS-INGLÊS	19
1.1. Sistema Educacional Brasileiro	23
1.2. Sistema Educacional Americano	25
2. FUNDAMENTAÇÃO TEÓRICA	28
2.1. A Linguística de <i>Corpus</i>	28
2.1.1. Definição de <i>Corpus</i> e tipos de <i>Corpora</i>	32
2.1.2. O Programa <i>WordSmith Tools</i> e suas Ferramentas	38
2.2. Lexicologia, Terminologia, Lexicografia e Terminografia	45
3. METODOLOGIA	55
3.1. A lista de palavras	57
3.2. Palavras-chave	59
3.3. As linhas de concordância	63
3.4. Processo de busca pelos correspondentes	64
4. ANÁLISE E DISCUSSÃO DOS DADOS	68
4.1. Análise e correspondência português-inglês.....	69
4.1.1. Educação.....	69
4.1.2. Ensino	71
4.1.3. Alunos.....	73
4.1.4. Escola/Colégio.....	74
4.1.5. Crianças	74
4.1.6. Vagas	75
4.1.7. Ano	75
4.1.8. Professor	75
4.1.9. Cursos	76
4.1.10. País.....	76
4.1.11. Aulas	76
4.1.12. Educadores.....	76
4.1.13. Sala.....	77
4.1.14. Matemática.....	77
4.1.15. Programa.....	77
4.1.16. Gestor.....	77
4.1.17. Gestão	77

4.1.18.	Aprendizagem.....	77
4.1.19.	Diversidade	77
4.1.20.	Renda	78
4.1.21.	Carreira	78
5.	PROPOSTA DE GLOSSÁRIO BILÍNGUE.....	80
5.1.	Apresentação inicial do Glossário Bilíngue de Educação	81
	CONSIDERAÇÕES FINAIS.....	84
	REFERÊNCIAS BIBLIOGRÁFICAS	86
	OBRAS DE REFERÊNCIA	90
	ANEXO 1 – Artigo do jornal <i>O Estadão</i>.....	91
	ANEXO 2 – Artigo do jornal <i>The New York Times</i>	92
	APÊNDICE A – Título dos artigos jornalísticos selecionados para este estudo.....	94
	APÊNDICE B – Substantivos em língua inglesa selecionados para a consulta de correspondência.....	103
	APÊNDICE C – Substantivos em língua portuguesa selecionados para compor a lista de termos e respectiva correspondência	105
	APÊNDICE D – Substantivos selecionados para análise lexical e busca pelos correspondentes	106
	APÊNDICE E – Substantivos encontrados no <i>Corpus</i> e com proposta de correspondência	107
	APÊNDICE F – Substantivos não encontrados no <i>Corpus</i> de estudo.....	108

INTRODUÇÃO

Educar é possibilitar ao ser humano acesso ao conhecimento, capaz de melhorar suas condições de existência em diversos aspectos. É possível constatar que nas últimas décadas houve grande ampliação da produção de conhecimento. Uma das razões foi o surgimento de diversas inovações tecnológicas que possibilitaram a universalização da informação. É possível saber o que se passa, quase que instantaneamente, em qualquer lugar do planeta. O jornal é um meio de fácil acesso e ampla circulação, especialmente na atualidade com o oferecimento da versão *online*, que pode ser consultada de qualquer lugar do mundo.

De um modo geral, a educação é tema comum e presente em toda sociedade, relaciona-se a diversos momentos da vida cotidiana individual e coletiva e está presente em muitas das informações divulgadas diariamente em textos jornalísticos e na mídia em geral.

Por essa razão, torna-se relevante conhecer e identificar a correspondência léxico-semântica de termos da educação comuns a toda e qualquer sociedade, para garantir a efetiva compreensão interlínguas.

Atentando-nos para a ausência de materiais terminológicos disponíveis na área da educação, especialmente no formato bilíngue, objetivamos, neste trabalho:

- 1) Identificar os termos com maior chavicidade na área de educação em língua portuguesa e língua inglesa a partir de *corpora* de artigos jornalísticos retirados de quatro jornais, sendo dois brasileiros (*O Globo* e *O Estadão*) e dois norte-americanos (*The New York Times* e *Los Angeles Times*).
- 2) Estabelecer a relação de correspondência léxico-semântica português-inglês dos termos estatisticamente significativos.

A partir desse estudo, objetivamos, especificamente:

- 1) Identificar os termos mais usados na área de Educação em língua portuguesa e língua inglesa.
- 2) Realizar uma investigação terminológica descritiva temática para estabelecer as relações de correspondência léxico-semântica na direção português-inglês;
- 3) Propor uma correspondência funcional para 42 termos português-inglês.

- 4) Propor a elaboração de um Glossário Bilingue hospedado em uma plataforma online.

Acreditamos que os resultados do presente estudo podem ser uma fonte para a elaboração de materiais terminológicos, como um glossário bilingue e poderá colaborar durante a tradução ou produção de textos por falantes de português.

A dissertação que ora apresentamos estrutura-se em cinco seções, excetuando a introdução, as considerações finais e as referências bibliográficas. Na primeira delas, discorreremos sobre a Educação como tema para estudo da correspondência léxico-semântica português-inglês, justificando o uso dos artigos jornalísticos como fonte para a busca dos termos. Apresentamos nessa seção, algumas características sobre os sistemas educacionais brasileiro e americano, contrastando e verificando as semelhanças e diferenças como base para estabelecer a correspondência interlínguas.

Na segunda seção, dissertamos acerca dos postulados teórico-metodológicos que sustentaram nossa pesquisa, isto é, a Linguística de *Corpus*, uma abordagem empírica para o estudo da língua que nos auxilia por meio do uso de *corpora* a investigar grandes quantidades de dados. Discorreremos também sobre os tipos de *corpora*, bem como os *corpora* utilizados nesta investigação e explicamos a ferramenta computacional *WordSmith Tool 6.0*, responsável pela extração dos dados.

Ainda nesta seção, abordamos as ciências do léxico, discorrendo sobre a Lexicologia e o estudo do léxico geral de uma língua, os materiais elaborados a partir do léxico geral, realizados pelos lexicógrafos, bem como o estudo da língua especializada, por meio da Terminologia. Abordamos a conceituação de palavra, vocábulo e termo e a utilização dos mesmos no âmbito de nossa pesquisa.

A terceira seção destina-se à descrição dos materiais e métodos utilizados para a obtenção dos dados de nossa pesquisa, no qual descrevemos o percurso desde a compilação dos *corpora*, criação de listas de palavras, lista de palavras-chave, concordância e procedimentos para o estabelecimento das correspondências português-inglês. Apresentamos a lista final com os substantivos selecionados para o estudo, com 42 termos da área de Educação (vide Apêndice D). Tendo em vista que utilizamos o *corpus* de estudo de língua inglesa como fonte para a busca de correspondência, quando o termo não foi encontrado em nosso *corpus*, foi indicada a ausência do termo. Dessa forma, propomos a correspondência léxico-semântica para 21 termos da língua portuguesa e apresentamos os termos não encontrados.

A seção seguinte apresenta a análise dos dados e a proposta de correspondência na direção português-inglês dos 42 termos simples da área de Educação, bem como as composições sintagmáticas (ensino fundamental e ensino médio, por exemplo). Utilizamos a lista das palavras-chave em língua portuguesa como a palavra-entrada para o verbete, tendo em vista que partimos da direção português-inglês, e a lista de palavras-chave em língua inglesa apoiou-nos na busca pela correspondência e quando o termo não se apresentava em nosso *corpus* de estudo, indicamos com a abreviatura NEC (não encontrado no *corpus*).

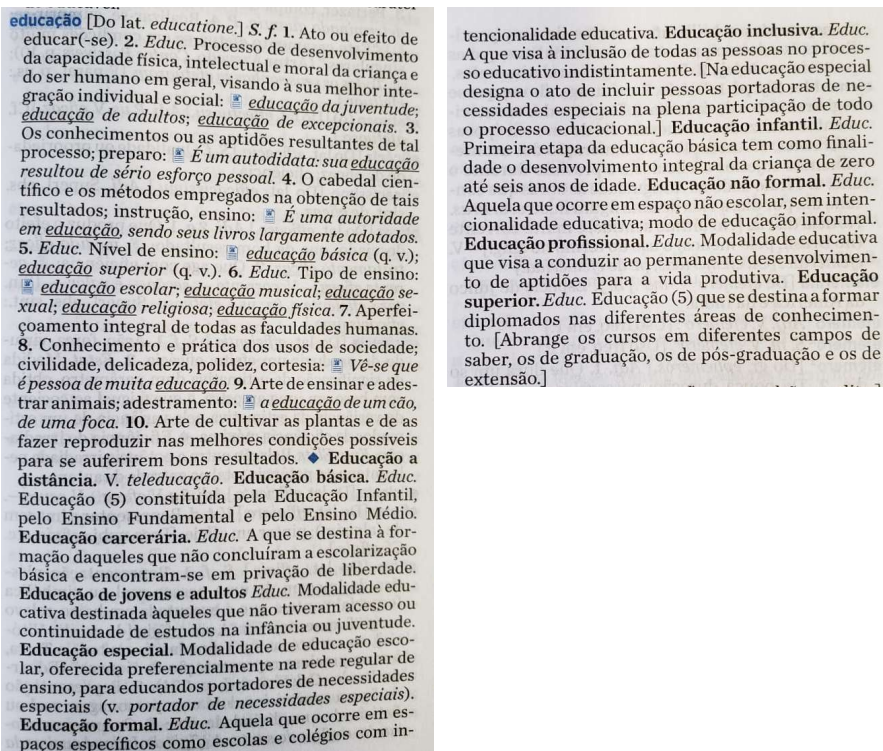
Na seção cinco, propomos a utilização dos dados obtidos na análise como subsídio para a elaboração de um glossário bilíngue de termos de Educação e sua hospedagem em um site. Nesta seção, exemplificamos a macroestrutura do glossário, bem como uma proposta de microestrutura.

Em seguida, discorreremos sobre as considerações finais em que avaliamos o estudo desenvolvido, bem como apresentamos os encaminhamentos futuros, seguidos das referências bibliográficas utilizadas e o apêndice.

1. A EDUCAÇÃO COMO TEMA PARA ESTUDO DA CORRESPONDÊNCIA LÉXICO-SEMÂNTICA PORTUGUÊS-INGLÊS

Em língua portuguesa, a abrangência de sentidos da palavra ‘educação’ perpassa, segundo o dicionário Aurélio (FERREIRA, 2010)¹, pelos sentidos: 1) ato ou efeito de educar-se; 2) processo de desenvolvimento da capacidade física, intelectual e moral da criança e do ser humano em geral, visando à sua melhor integração individual e social; 3) os conhecimentos ou as aptidões resultantes de tal processo; 4) o cabedal científico e os métodos empregados na obtenção de tais resultados; instrução, ensino; 5) nível de ensino (educação básica, educação superior); 6) tipo de ensino (educação física, educação musical, educação religiosa, educação sexual); 7) aperfeiçoamento integral de todas as faculdades humanas; 8) conhecimento e prática dos usos de sociedade; civilidade, delicadeza, polidez, cortesia; 9) arte de ensinar e adestrar animais; 10) arte de cultivar as plantas; além dos termos complexos como: Educação básica, Educação de jovens e adultos, Educação inclusiva, entre outros, conforme verbete na figura 1:

Figura 1 - Verbetes 'educação' no dicionário Aurélio



Fonte: Dicionário Aurélio (FERREIRA, 2010)

¹ Devido a representatividade da língua portuguesa, escolhemos o Dicionário Aurélio (FERREIRA, 2010) para este estudo.

Em língua inglesa, o *Longman Dictionary of Contemporary English (LDCE)* (LONGMAN, 1987), apresenta um verbete com uma microestrutura mais concisa, sem muitos exemplos, apenas com uma acepção em que diz: o processo pelo qual o caráter e mente de uma pessoa são desenvolvidos através do ensino, por meio de instrução formal em escola ou universidade. Há também quatro exemplos: dois referindo-se à **educação** escolar ('instituto de educação' e [...] completou sua **educação** na Suíça.) e dois que abrangem o termo das seguintes formas ('Ministro da **Educação**' e 'O governo está gastando muito dinheiro em **educação**').² Além disso, direciona o consulente a procurar por '*Higher education*', '*Further education*' e '*Teach*', conforme figura 2:

Figura 2 - Verbetes 'education' no dicionário LDCE

ed-u-ca-tion /,edʒ'keɪʃən/ n [S;U] the process by which a person's mind and character are developed through teaching, esp. through formal instruction at a school or college: *an institute of adult education/the Minister of Education/She completed her education in Switzerland/The government is spending a lot of money on education.* — see also FURTHER EDUCATION, HIGHER EDUCATION; see TEACH (USAGE)

Fonte: *Longman Dictionary of Contemporary English* (LONGMAN, 1987)

O Dicionário Aurélio (FERREIRA, 2010) traz dez acepções da palavra **educação**, além dos termos complexos apresentados no final do verbete. Já o LDCE apresenta uma acepção, quatro sentenças e o direcionamento para outros verbetes que possam trazer mais informações acerca do tema. A acepção apresentada no LDCE é restrita ao ambiente escolar, enquanto o Dicionário Aurélio (FERREIRA, 2010) dispõe de outras acepções, como em 8 (Figura 1) cortesia e polidez.

De um modo geral, a educação é tema comum e presente em toda sociedade, relaciona-se a diversos momentos da vida cotidiana individual e coletiva e está presente em muitas das informações divulgadas diariamente em textos jornalísticos e na mídia em geral.

Ocorre que, ao deparar-se com a palavra **educação**, cada indivíduo a compreenderá segundo a sua história e contexto de uso. Podemos dizer que um professor, certamente, pensará nas suas aulas, seus alunos, no ensino básico ou superior ou nas políticas públicas de educação; já o cidadão que exerça outra função não relacionada à escola e ao ensino pode fazer associação com o comportamento adequado das pessoas, com a sua vivência familiar ou com a sua escolarização. De qualquer forma, a

² Traduções e grifos nossos.

compreensão individual estará sempre relacionada com o senso comum e uso da língua de cada região ou país.

No mundo atual, cada vez mais globalizado, é indispensável a comunicação para além das fronteiras. Com o advento da Internet, é cada vez mais fácil e usual o acesso ao conhecimento e informações na mídia de outros países e em outras línguas. Decorre das questões apresentadas a relevância de conhecer e identificar a correspondência léxico-semântica de termos comuns da área da Educação nos contextos brasileiro e americano, para garantir a efetiva compreensão interlínguas e para garantir a compreensão de informações e troca de saberes. Sendo assim, este trabalho apresentará um estudo da correspondência português-inglês de 42 termos da área da Educação a partir de um *corpus* de artigos jornalísticos.

Para exemplificar, destacamos a palavra ‘educação’, nos seguintes contextos, a fim de que contrastemos as possibilidades abordadas com o tema ‘educação’ nos artigos jornalísticos do par português-inglês, o primeiro artigo, escrito em língua portuguesa, intitula-se ‘Reforma do Ensino Médio exige atenção para evitar aumento das desigualdades’ (ALTENFELDER, 2017)³, o qual destacamos a palavra **educação** para verificarmos o contexto de uso:

Em uma sociedade desigual como a nossa, políticas educacionais devem se pautar pela democratização do direito à educação⁴ e pela busca de equidade. No caso do Ensino Médio, uma adequada política deve, portanto, buscar garantir que todos os alunos possam ter as mesmas possibilidades de desenvolvimento pessoal, de inserção no mercado de trabalho e de acesso ao Ensino Superior, independentemente de estudarem em tempo de integral ou parcial, ou em turmas do diurno ou noturno. Garantir a democratização do direito à educação e a equidade é um grande desafio. A nova legislação do ensino médio, porém, contém medidas que podem dificultar avanços nessa direção. Uma delas é a implementação paulatina de escolas de tempo integral. Sabemos – e pesquisas mostram isto – que escolas desse tipo oferecem uma educação de maior qualidade. (ALTENFELDER, 2017)

Observamos que **educação** nesse contexto se encaixa nas acepções 5) nível de ensino (educação básica, educação superior) e 6) tipo de ensino (educação física, educação musical, educação religiosa, educação sexual) do Dicionário Aurélio (FERREIRA, 2010), abordando questões do âmbito escolar e de um nível específico de

³ O artigo completo encontra-se no Anexo 1.

⁴ Todos os grifos apresentados no artigo são de nossa autoria.

educação. Pelo tema abordado, há grande possibilidade de encontrarmos outros termos relacionados à **educação** nesse contexto.

Em inglês, apresentamos o artigo intitulado como ‘*Crime in New York City Schools is at a Record Low, City says*’ (HARRIS, 2017)⁵:

[...] The former United States secretary of education John B. King Jr., who was previously New York State’s education⁶ commissioner, said that the data “rarely reflects the realities of school health and safety.” Mark Rampersant, the deputy chief executive for safety and security at the city’s Education Department, said Tuesday that the state system counted an episode involving a first-grader hitting a classmate with an empty milk carton as assault with a weapon. The state recently adopted a new system for tracking school violence, which went into effect in July. (HARRIS, 2017)⁷

Observa-se que o uso da palavra **educação** aproxima-se do contexto de uso do verbete apresentado no LDCE, como ‘secretário de educação’.

Sendo assim, nota-se que os artigos jornalísticos apresentam textos em que **educação** se relaciona às acepções apresentadas nos dicionários de língua portuguesa e língua inglesa, sendo uma fonte viável para o levantamento dos termos relacionados à área.

Escolhemos o tema **educação**, porque o mesmo se faz presente na vida de todas as pessoas, direta ou indiretamente, e porque é nossa intenção colaborar para a efetiva compreensão das palavras decorrentes e de seus correspondentes em ambas as línguas. Propomos, assim, o levantamento das palavras mais frequentes da área de Educação a partir de artigos jornalísticos e apresentamos a correspondência interlínguas, bem como a proposta de elaboração de um glossário bilíngue da área em estudo e a hospedagem em um site.

Ao tratar do tema **educação** e para embasar a análise, torna-se essencial conhecer as semelhanças e diferenças entre o sistema educacional brasileiro e o sistema

⁵ O artigo completo encontra no Anexo 2.

⁶ Todos os grifos presentes no artigo são de nossa autoria.

⁷ O ex-secretário de educação dos Estados Unidos, John B. King Jr., que anteriormente era comissário de educação do Estado de Nova York, disse que os dados “raramente refletem as realidades da saúde e segurança escolar”. Mark Rampersant, vice-presidente executivo de segurança no Departamento de Educação da cidade, disse na terça-feira que o sistema estadual contou um episódio envolvendo uma aluna da primeira série que bateu em um colega de classe com uma embalagem de leite vazia como um assalto com arma. O estado adotou recentemente um novo sistema de rastreamento da violência escolar, que entrou em vigor em julho. (tradução nossa)

educacional americano, uma vez que serão referência para sentidos comuns e específicos em cada país.

Abordaremos definições básicas do sistema educacional brasileiro no subitem 1.1 e as definições básicas do sistema educacional americano no subitem 1.2.

1.1.Sistema Educacional Brasileiro

A educação no Brasil é organizada segundo a Lei de nº 9.394 de Diretrizes e Bases da Educação Nacional, de 20 de dezembro de 1996 (LDB 9.394/96) incumbida por definir os órgãos administrativos responsáveis e organizar os níveis e modalidades de ensino, regularizando o sistema educacional brasileiro com base nos princípios presentes na Constituição. Os órgãos responsáveis pela educação, em nível federal, são o Ministério da Educação (MEC) e o Conselho Nacional de Educação (CNE). Em nível estadual, temos a Secretaria Estadual de Educação (SEE), o Conselho Estadual de Educação (CEE), a Delegacia Regional de Educação (DRE) ou Subsecretaria de Educação. E, por fim, em nível municipal, existem a Secretaria Municipal de Educação (SME) e o Conselho Municipal de Educação (CME).

Atualmente a estrutura educacional regular compreende a educação básica, composta pela educação infantil, ensino fundamental e ensino médio. Além dessas três etapas, segundo o artigo 21 da Lei n.º 9.394/96, compreende-se ainda o nível superior e outras quatro modalidades de ensino: a) Educação de jovens e adultos (ensino fundamental e médio); b) Educação profissional ou técnica; c) Educação especial; e d) Educação a distância (EAD).

A educação infantil, primeira etapa da educação básica, é oferecida em creches, para crianças de até 3 anos de idade e em pré-escolas, para crianças de 4 a 5 anos. O ensino fundamental, com duração mínima de nove anos, é obrigatório e gratuito na escola pública. É dividido em duas partes, por vezes chamado de fundamental ciclo 1 e fundamental ciclo 2. O ensino médio, etapa final da educação básica, tem duração mínima de três anos e atende a formação geral do educando, podendo incluir programas de preparação geral para o trabalho e, facultativamente, a habilitação profissional.

Resumindo, o sistema educacional brasileiro é composto pela: a) educação infantil (creche e pré-escola); b) Ensino Fundamental (divido em 2 partes); e c) Ensino Médio, conforme Quadro 1:

Quadro 1 - Estrutura do sistema educacional brasileiro

Etapa	Nomenclatura
Educação Infantil	Creche
	Pré-Escola
Ensino Fundamental	(1º ano)
	(2º ano)
	(3º ano)
	(4º ano)
	(5º ano)
	(6º ano)
	(7º ano)
	(8º ano)
	(9º ano)
Ensino Médio	1º ano
	2º ano
	3º ano
Ensino Superior	

Fonte: Quadro adaptado de Bastos (2017).

Conforme Quadro 1, destaca-se as três etapas da educação brasileira obrigatórias e de direito a todos os cidadãos: Educação Infantil, obrigatória a partir da pré-escola, Ensino Fundamental e Ensino Médio. O Ensino Superior é oferecido tanto pelos governos estaduais e federais, quanto pelo setor privado. Como demonstrado no quadro, há diferentes formas convencionais na língua portuguesa, sendo que os termos complexos variam entre as construções utilizando ‘ensino’ ou ‘educação’.

O Ensino Superior brasileiro, oferecido tanto pelos governos estaduais e federais ou setor privado, utilizam o mesmo sistema de ingresso. O primeiro critério é a conclusão do Ensino Médio, candidatar-se a uma universidade e realizar uma prova objetiva que abrange áreas de humanas, exatas e biológicas. A classificação é feita com base na pontuação dos candidatos.

Concomitante ao ensino médio ou após sua conclusão, há a oferta de cursos técnico ou profissionalizantes, oferecidos tanto pelo setor privado quanto pelo setor público.

Dessa forma, organiza-se o sistema educacional brasileiro bem como suas etapas obrigatórias e de direito às crianças e jovens. A próxima seção (1.2) dedica-se a sistematização do sistema educacional americano, a fim de que tenhamos um parâmetro dos pares de línguas estudados nesta investigação.

1.2.Sistema Educacional Americano

A educação nos Estados Unidos também é controlada pelos níveis federal, estadual e local. Os alunos podem frequentar escolas públicas, privadas ou domésticas⁸. Pode-se dizer que a organização ocorre em três fases (*elementary, middle school e high school*)

O ensino obrigatório, de modo geral, varia de estado para estado e começa a partir do 1º ano do nível '*elementary*'.

⁸ O ensino doméstico é a instrução progressiva de um estudante sob responsabilidade dos pais ou responsável. É uma prática comum e permitida nos Estados Unidos. Os pais que optam por esse tipo de ensino devem enviar uma carta ao superintendente do distrito escolar, com as informações pessoais do aluno e dos pais. Devem manter portfólios e cadernos para possíveis fiscalizações e realizar exames que comprovem a evolução do aprendiz. Para obtenção do certificado de conclusão do ensino médio, o aluno pode realizar o SAT (equivalente ao ENEM, no Brasil) e obter o documento que comprova sua escolarização. (XIMENES, 2018; RAY, 2017)

Quadro 2 - Estrutura do sistema educacional americano

Etapa	Nomenclatura
Pre-Kindergarten	
Lower Elementary	Kindergarten 1st grade 2st grade
Upper Elementary	3 rd grade 4 th grade 5 th grade
Middle School	6 th grade 7 th grade 8 th grade
High School	9th grade = <i>freshman</i> 10th grade = <i>sophomore</i> 11th grade = <i>junior</i> 12th grade = <i>senior</i>
Higher education	

Fonte: Quadro elaborado pela autora, adaptado de Schütz & Kanomata (2011)

Conforme Quadro 2, há cinco etapas educacionais antes da *Higher education* (Ensino Superior), sendo obrigatório a partir do primeiro ano da *Lower Elementary*. A principal característica que visualizamos no quadro refere-se a *High school* (ensino Médio) que possui a nomenclatura de *9th a 12th grades*, e cada série/ano possui um termo específico *freshman, sophomore, junior e senior*.

Já o ingresso nas universidades americanas desenvolve-se de outra maneira. Diferente do sistema brasileiro, o desempenho dos alunos durante o ensino médio, bem como as atividades extracurriculares são requisitos para a aprovação do aluno na universidade. Além disso, as universidades pedem cartas de recomendações, escritas pelos professores do aluno/candidato, comprovando o bom desempenho nos anos escolares.

O processo começa com a “*application*” (inscrição) que é o envio da documentação exigida, como: histórico escolar (*transcripts*), cartas de recomendação (*recommendation letter*) e uma redação (*essay*) escrita pelo próprio candidato.

Algumas universidades exigem uma prova escrita, chamada “*SAT*” (*Scholastic Aptitude Test*), que avalia habilidades de raciocínio em matemática, interpretação de texto e escrita em inglês; ou o “*ACT*” (*American College Testing*), composto por quatro seções:

inglês, matemática, interpretação de texto e raciocínio científico, além de uma seção optativa de escrita. Existe ainda o *SAT Subject Test (ou SAT II)*, que serve para avaliar o conhecimento em uma área específica. Além disso, as universidades podem realizar entrevistas com os candidatos.

Em nossa proposta de correspondência dos termos da Educação português-inglês, tomamos esse conhecimento como base de análise e buscamos abordar essas informações de forma concisa e de fácil acesso ao consulente, sempre levando em consideração que quando escrevemos ou traduzimos algo devemos fazer uso do que é convencional na língua.

A próxima seção destina-se a descrição dos princípios teóricos que deram sustentação a esta pesquisa e nos conduziram aos resultados.

2. FUNDAMENTAÇÃO TEÓRICA

Nesta parte, apresentaremos as principais teorias que deram sustentação ao nosso estudo, em especial, no que se refere aos Fundamentos da Linguística de Corpus (BERBER SARDINHA, 2000, 2004; TOGNINI-BONELLI, 2001, BIBER, 1998, SINCLAIR, 1991; BEVILACQUA, 2013), a alguns princípios da Terminologia e Terminografia (AUBERT, 1998; BARROS, 2004; KRIEGER & FINATTO, 2004, BARBOSA, 1999) e pressupostos da Lexicologia e Lexicografia (BIDERMAN, 1987, 1996, 1998, 2001; WELKER, 2004)

2.1.A Linguística de *Corpus*

Para tornar possível o manuseio de uma grande quantidade de dados, faz-se necessária a utilização de um aparato teórico-metodológico que nos permita a extração semiautomática dos dados que serão objetos de análise nesta pesquisa, a realização desta tarefa sem a utilização de uma máquina e de ferramentas apropriadas torna-se inviável. Sendo assim, nossa pesquisa tem como base a Linguística de *Corpus*, que assim como destaca Bevilacqua (2013) estabelece os princípios e critérios para a compilação dos *corpora* e auxilia nas etapas metodológicas terminográficas:

[...] além de estabelecer os princípios e critérios para a compilação de corpora também oferece recursos e ferramentas que auxiliam nas diferentes etapas metodológicas terminográficas: desde a própria compilação de corpora, passando pela identificação de candidatos a termos e fraseologias e chegando à identificação de elementos que permitem a elaboração de definição. (BEVILACQUA, 2013, p. 17).

Apesar do termo Linguística de *Corpus* (doravante LC) ser uma designação moderna, sua história é antiga. Aston (2013, p. 1) atribui as primeiras aparições do que se pode ser considerado LC, no século 13, quando Hugh of St. Cher empregou 500 monges para listar todas as palavras da Bíblia. No século XIX, Preyer investigava as falas das crianças por meio de diários registrados pelos pais. Em 1897, Kading utilizou um *corpus* de 11 milhões de palavras para investigar a ortografia do alemão. Um século depois, nos anos de 1940 e 1947, foram realizados estudos com *corpora* em pedagogia da língua estrangeira, iniciando a utilização de listas de vocabulário feitas a partir de *corpus* para o ensino de língua estrangeira. (MCENERY & WILSON, 2001, p. 3). O trabalho desenvolvido nessa época é digno de admiração, uma vez que eram criados *corpora* e

desenvolvidos métodos para análise de dados sem qualquer ferramenta ou suporte tecnológico.

Porém, conforme nos apresenta Berber Sardinha (2004) o desenvolvimento da LC desacelerou por volta da década de 1950, período que coincide com o aparecimento do *Syntactic Structures de Chomsky*, publicado em 1957. A teoria chomskyana provocou uma mudança de paradigma nos estudos linguísticos da época e motivou o surgimento de outros. Dentre os postulados de Chomsky (1957), estava a suspeição das pesquisas de cunho empírico. As investigações baseadas em grandes quantidades de dados foram vistas como perda de tempo e dinheiro. Na visão do autor, todo e qualquer pesquisador deveria confiar apenas em sua intuição e no que é gramaticalmente correto. O linguista deveria fazer um modelo da competência (conhecimento internalizado) e não do desempenho linguístico (evidência externalizada). Assim, a linguística possuía uma visão racionalista, fundamentada na introspecção, observação e intuição do linguista.

Contrária à visão racionalista, a LC possui uma visão da linguagem como sistema probabilístico e uma abordagem empirista, que parte da observação de dados coletados para determinado estudo. O cerne da LC está “no desempenho, na observação da evidência, na linguagem atestada por meio de exemplos reais extraídos de um *corpus*”. (NAVARRO, 2011, p. 16)

Aliada à dificuldade da compilação de corpora manualmente, em que demandava muito tempo e uma equipe grande, juntamente com a possibilidade de falha humana, tendo em vista que contar palavras torna-se uma tarefa tediosa e a probabilidade de erro é grande. Culminando com os ideais de Chomsky (1957), a pesquisa empírica foi perdendo sua popularidade.

A partir dos anos 1980, com a introdução dos computadores pessoais, o cenário muda novamente e a pesquisa linguística baseada em *corpora* ganha impulso. O impacto causado pela informática na LC é comparado ao papel do microscópio e do telescópio nas ciências, que permitiram que os cientistas observassem coisas que jamais tinham visto (STUBBS, 1996). Com esta ferramenta, há a possibilidade de armazenamento de grande quantidade de dados digitalizados cuja automaticidade da manipulação contribuiu para o avanço da LC.

O uso do computador, além de acelerar e intensificar o manuseio de grandes quantidades de dados e permitir maior armazenamento, pôde colocar a matemática a favor da linguística, fazendo com que cálculos simples e complexos pudessem ser realizados rapidamente e sem margens para erros. A estatística passou a ser incorporada à LC e a

conceder-lhe maior teor científico. (VASILÉVSKI, 2007 apud KADER e RICHTER, 2013)

Com a tecnologia e a matemática a seu favor, a LC tem avançado dia após dia, abrangendo diferentes áreas de aplicação e servindo de apoio aos pesquisadores. Conforme relata Berber Sardinha (2004), “[...] na Grã-Bretanha, um dos centros mais desenvolvidos de pesquisa várias universidades dedicam-se às pesquisas baseadas em *corpus* para a descrição de diferentes aspectos relacionados à linguagem.” (BERBER SARDINHA, 2004, p. 5). Nos países escandinavos (Noruega, Suécia e Dinamarca) há centros dedicados à pesquisa em LC há vários anos.

Nos Estados Unidos, apesar da facilidade em obter recursos de informática, a LC desenvolveu-se de maneira mais modesta, em função da influência da linguística gerativa transformacional, ramo forte entre os centros de pesquisa em linguística. Por outro lado, os Estados Unidos desenvolveram trabalhos mais expressivos na área da Pesquisa em Processamento de Linguagem Natural, presente tanto em nível acadêmico como nas pesquisas com fins comerciais. (BERBER SARDINHA, 2004, p. 5)

No Brasil, a pesquisa em *corpus* acontece em centros voltados ao Processamento de Linguagem Natural, Lexicografia e Linguística Computacional. A citar, o Programa de Pós-Graduação em Linguística Aplicada (LAEL) da PUC-SP, os projetos da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, COMET (*Corpus* Multilíngue para Ensino e Tradução), CorTec (*corpora* técnicos em inglês e português), CoMAprend (*corpora* de aprendizes) e os grupos de pesquisas liderados pelos docentes do Programa de Estudos Linguísticos da UNESP, campus de São José do Rio Preto.

Os recursos da LC, suas ferramentas e utilitários têm contribuído para avanços no campo dos estudos lexicológicos e terminológicos, não só pelo fato de permitirem a organização e seleção precisa dos dados, mas por propiciar investigações assertivas com diferentes finalidades (SANTOS, 2015). Dentre estas, a literatura e o ensino de línguas, tanto na modalidade descritiva, discorrendo sobre padrões da linguagem, combinações e fraseologias, quanto na área aplicada, como auxílio para as áreas de lexicografia e terminografia e produção de materiais didáticos, como dicionários, glossários e vocabulários.

A LC, como assumimos nesta pesquisa, pode ser vista como aparato teórico-metodológico para as investigações. Tonigni-Bonelli (2001) não a considera uma simples metodologia ou conjunto de regras aplicadas a situações específicas. A autora refere-se à LC como *pre-application methodology*, visto que define suas próprias regras e

conhecimentos teóricos antes de serem aplicados, possibilitando ao pesquisador a escolha de novos parâmetros para dar conta de explicar e descrever os dados. (TOGNINI-BONELLI, 2001)

Para justificar esse viés teórico da LC, a autora afirma que há dois tipos de pesquisas possíveis de serem desenvolvidas por meio da LC, a pesquisa **baseada em corpus** (*corpus-based approach*) ou a pesquisa **orientada pelo corpus** (*corpus-driven approach*). A abordagem baseada em *corpus* representa o uso da LC como uma ferramenta metodológica, em que o *corpus* funciona como uma fonte de exemplos destinado a atestar determinado postulado teórico. Sendo que na abordagem dirigida pelo *corpus*, os postulados teóricos só podem ser formulados com base nas evidências do *corpus*, o que pode ser constatado nas pesquisas de Sinclair (1991), por exemplo. (TOGNINI-BONELLI, 2001)

Assim como a autora afirma que a abordagem baseada em corpus funciona como uma fonte de exemplos, que nada mais são do que evidências para determinado estudo, Sinclair (1991) ressalta a necessidade da LC estar acompanhada da língua em uso, do contexto, da situação de uso, afirmando que o estudo compreensivo da língua deve ser baseado em evidências textuais. Não há possibilidade de retirar dados de outro lugar senão da língua em uso, assim como “não se estuda botânica por meio de flores artificiais”⁹ (SINCLAIR, 1991, p. 6).

Constatar e apoiar-se em evidências é assumir a língua como sistema probabilístico, assim como relata Berber Sardinha (2004) ao retomar Halliday (1991), em que a probabilidade de ocorrer determinados fenômenos em contextos específicos é grande e não casual. Da mesma maneira, Sinclair (1996) ressalta a não aleatoriedade da língua, afirmando que tudo é planejado, determinado e organizado:

O estudo de corpus permite-nos observar os usos que as pessoas fazem da língua, reuni-los, e olhá-los de uma maneira científica. Quando fazemos isso, notamos que há uma grande quantidade de coocorrência de palavras que não são aleatórias, mas claramente regulares e frequentes. De fato, a ideia de qualquer tipo de aleatoriedade, de qualquer tipo de ocorrência determinada pelo acaso é bastante remota na língua. Tudo é altamente determinado, altamente planejado e altamente organizado. (SINCLAIR, 1996, p. 99)¹⁰

⁹ *One does not study all of botany by making artificial flowers.*

¹⁰ *The study of corpus texts allow us to stand back from peoples' usage, gather it together, and look at it in a scientific fashion. When you do this, you begin to notice that there is a great deal of co-occurrence of words that are not all random but, quite clearly, regular and frequent. In fact, the idea of any kind of randomness, any kind of chance occurrence in language is very remote indeed. It is all highly determined, highly planned, highly organised.*

Assim, nos aproximamos do pensamento de Tognini-Bonelli (2001) e reconhecemos a natureza teórico-metodológica da LC, devido ao fato de ser esta, ora uma fonte, ora um instrumento para pesquisa.

O emprego dos *corpora*, seja paralelo ou comparável revela-se também na metodologia para a identificação de equivalentes tradutórios (Tognini-Bonelli, 1996, 2001) e (Tognini-Bonelli & Manca, 2004), que evidencia as relações estabelecidas pelos itens lexicais em contextos típicos. Segundo Tognini-Bonelli (1996, p. 200), retomando Firth (1957):

[...] grande parte do significado de um item surge das relações com o seu cotexto imediato, bem como do contexto geral da situação. A observação e a análise das características formais presentes no cotexto de um item linguístico suscitam descobertas sobre seu significado e sobre sua função. Esta é, na minha opinião, onde a teoria firthiana do significado pode ser utilizada no âmbito da Linguística de Corpus: observações e generalizações obtidas do cotexto são deslindáveis e correspondências com significado podem, portanto, ser extraídas com consistência.¹¹

Passamos à conceituação de *corpus*, bem como os tipos disponíveis e as características dos *corpora* dessa pesquisa.

2.1.1. Definição de *Corpus* e tipos de *Corpora*

Antes mesmo do advento da LC, investigações linguísticas já utilizavam *corpora* para a obtenção de seus resultados. Referidas por McEnery e Wilson (2001) como linguística de corpus preliminares (*early corpus linguistics*), destacam-se as investigações sobre as línguas indígenas, conduzidas por Boas (1940); trabalhos sobre aquisição da linguagem, de Stern (1924), outros voltados à pedagogia da língua, conduzidos por Thorndike (1921) e Palmer (1933), assim como pesquisas na área de sintaxe e semântica.

O uso de *corpora* nas pesquisas, pode ser considerado como uma abordagem empírica que segue procedimentos, que resultam em um postulado teórico. Assim como esclarece Tognini-Bonelli (2001):

O trabalho com corpus pode ser visto como uma abordagem empírica que, similarmente a todo tipo de indagação científica, tem como ponto

¹¹ [...] a major part of the meaning of an item arises from its relations with its immediate co-text as well as the general context of situation. The observation and analysis of the formal features present in the co-text of a linguistic item, therefore, is taken to yield insights into its meaning and function. This is, in my opinion, where the Firthian theory of meaning can be used in the framework of Corpus Linguistics: observations and generalizations drawn from the co-text are inherently tractable, and correspondences with meaning can therefore be consistently drawn.

de partida dados autênticos. O procedimento para descrição dos dados que constituem um corpus é, portanto, indutivo, haja vista que as afirmações de natureza teórica sobre a língua e a cultura decorrem da observação de dados reais. A observação dos fatos da língua leva à formulação de uma hipótese que dê conta desses fatos; o que, por sua vez, conduz à uma generalização baseada na evidência de padrões recorrentes nas linhas de concordância; a última etapa é a unificação dessas afirmações em um postulado teórico (TOGNINI-BONELLI, 2001, p. 2)¹²

Há inúmeras definições de *corpus* presentes na literatura. Tognini-Bonelli (2001) destaca corpus como sendo

[...] uma coleção de textos representativos de uma dada língua, reunidos de forma que possam ser utilizados para análise linguística. Em geral, parte-se do pressuposto que os dados armazenados sejam ocorrências naturais, coletados de acordo com critérios explícitos e com um propósito em mente, representando grandes blocos da língua selecionados em conformidade com uma tipologia específica (TOGNINI-BONELLI, 2001, p. 2).¹³

Sánchez (1995) partilha da mesma ideia, porém incorpora em sua definição a necessidade do uso do computador para o processamento dos dados, sendo assim, relata de forma mais abrangente as principais características de um *corpus* eletrônico:

Um conjunto de dados linguísticos (pertencentes ao uso geral 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. (SÁNCHEZ, 1995, p.8-9; traduzido por Berber Sardinha, 2004, p.18).

Essa definição, para o autor, reúne os pontos mais importantes a serem seguidos na construção de um corpus:

- a) **A origem:** os dados devem ser autênticos.

¹² *Corpus work can be seen as an empirical approach in that, like all types of scientific enquiry, the starting point is actual authentic data. The procedure to describe the data that makes use of a corpus is therefore inductive in that it is statements of a theoretical nature about the language or the culture which are arrived at from observations of the actual instances. The observation of language facts leads to the formulation of a hypothesis to account for these facts; this in turn leads to a generalization based on the evidence of the repeated patterns in the concordance; the last step is the unification of these observations in a theoretical statement*

¹³ *A corpus can be defined as a collection of texts assumed to be representative of a given language put together so it can be used for linguistic analysis. Usually that assumption is that the language stored in a corpus is naturally occurring, that it is gathered according to explicit design criteria, with a specific purpose in mind, and with a claim to represent larger chunks of language selected according to a specific typology.*

- b) **O propósito:** o corpus deve ter a finalidade de ser um objeto de estudo linguístico.
- c) **A composição:** o conteúdo do corpus deve ser criteriosamente escolhido.
- d) **A formação:** os dados do corpus devem ser legíveis por computador.
- e) **A extensão:** o corpus deve ser vasto para ser representativo. (BERBER SARDINHA, 2004, p.18-19)

Tagnin (2015) reafirma a importância do cuidado na seleção dos textos que serão utilizados na composição do *corpus*, para que os resultados provenientes da pesquisa sejam de fato confiáveis:

Os corpora devem ser criteriosamente construídos, de acordo com o objetivo a que se destinam, isto é, deve-se ter a certeza de que os textos compilados são representativos do campo que se deseja pesquisar e são de fonte confiável, caso contrário os resultados podem não ser. (TAGNIN, 2015, p. 20)

Todas as definições mencionadas ressaltam a importância de considerar o propósito da pesquisa para que o *corpus* seja construído para a finalidade adequada e consiga alcançar o resultado proposto/esperado. Os corpora podem abarcar as mais diversas áreas e possuir diferentes propósitos, para tanto precisam ser adequados à investigação a que se propõe, “para serem adequados, os corpora devem ser afinados com os objetivos da análise”. (HASAN, 1992 apud BERBER SARDINHA, 2004, p. 29).

A depender do objetivo da pesquisa, vários tipos de *corpora* podem ser empregados. Segundo Berber Sardinha (2000) há quatro pré-requisitos a serem seguidos para a formação de um *corpus* computadorizado, são eles: 1) o *corpus* deve ser composto por textos autênticos, em linguagem natural. Ou seja, não podem ter sido produzidos com a finalidade de ser alvo de pesquisa; 2) os textos devem ser autênticos, no sentido de que sejam produzidos por falantes nativos da língua. Caso contrário, devem ser qualificados como *corpora* ‘de aprendizes’; 3) o conteúdo do *corpus* deve ser escolhido criteriosamente, seguindo algumas regras propostas pelo próprio criador, a fim de que alcance os objetivos esperados e; 4) representatividade do *corpus*, que se caracteriza entre outras coisas, quanto a extensão do *corpus*.

Considerados os pré-requisitos para a formação de um corpus computadorizado, é imprescindível seguir alguns critérios para compilar um *corpus* de estudo. Esses critérios auxiliam na compilação de um *corpus* adequado, segundo Berber Sardinha (2004). Um *corpus* pode ser, no que tange ao modo de comunicação, ***corpora de língua***

falada, compostos por porções de fala transcritas ou **corpora de língua escrita**, compostos por textos escritos, impressos ou não. Pode compreender um período de tempo (**sincrônico**); vários períodos (**diacrônico**); representar o tempo atual ou representar um período passado (**contemporâneo** ou **histórico**)

No que tange a seleção dos dados, os *corpora* podem ser de **amostragem**, compostos por porções de texto ou de variedades textuais, planejado para ser uma amostra finita da linguagem como um todo; *corpus monitor*, em que a composição é reciclada para refletir o estado atual de uma língua, a exemplo do *Corpus of Contemporary American English (COCA)*; **dinâmico ou orgânico**: o crescimento e diminuição são permitidos, este tipo qualifica o *corpus* monitor; **estático**: oposto ao dinâmico, ele caracteriza o *corpus* de amostragem e pode ser **equilibrado**, quanto aos componentes (gêneros, textos etc.) que são distribuídos em quantidades semelhantes (por exemplo, mesmo número de textos por gênero).

Dependendo de sua autoria, o *corpus* pode ser classificado como **corpus de aprendiz**, composto por textos escritos por não nativos da língua, geralmente utilizados por estudantes de uma língua estrangeira e utilizados para a descrição da linguagem utilizada ou auxílio para produção de materiais didáticos que funcionem para suprir possíveis lacunas no aprendizado. Os **corpora de estudo**, são aqueles compilados para o desenvolvimento de determinado estudo, aquele que se pretende descrever. Os **corpora de referência**, são utilizados para fins de contraste com o *corpus* de estudo

Em relação ao conteúdo do *corpus*, há os **corpora especializados**, nos quais textos específicos compõem o *corpus*; podem ser **monolíngues** e compostos por apenas uma língua; ou **multilíngues**. (BERBER SARDINHA, 2004)

Dessa forma, os critérios para compilação dos *corpora*, propostos por Berber Sardinha (2004) seriam:

Quadro 3 - Critérios para compilação dos *corpora*

Modo	Falado: composto de porções de fala transcritas. Escrito: composto de textos escritos, impressos ou não.
Tempo	Sincrônico: compreende um período de tempo. Diacrônico: compreende vários períodos de tempo. Contemporâneo: representa o período de tempo corrente. Histórico: representa um período de tempo passado.
Seleção	De amostragem (sample corpus): composto por porções de texto ou de variedades textuais, planejado para ser uma amostra finita da linguagem como um todo. Monitor: a composição é reciclada para refletir o estado atual de uma língua; opõe-se a corpora de amostragem. Dinâmico ou orgânico: o crescimento e diminuição são permitidos; qualifica o corpus monitor. Estático: oposto a dinâmico; caracteriza o corpus de amostragem. Equilibrado (balanced): os componentes (gêneros, textos etc.) são distribuídos em quantidades semelhantes (por exemplo, mesmo número de textos por gênero).
Conteúdo	Especializado: os textos são de tipos específicos (gênero ou registros definidos). Regional ou dialetal: os textos são provenientes de uma ou mais variedades sociolinguísticas específicas. Multilíngue: inclui idiomas diferentes.
Autoria	De aprendiz: os autores dos textos não são falantes nativos. De língua nativa: os autores são falantes nativos.
Finalidade	De estudo: o corpus que se pretende descrever. De referência: usado para fins de contraste com o corpus de estudo. De treinamento ou teste: construído para permitir o desenvolvimento de aplicações e ferramentas de análise.

Fonte: Berber Sardinha (2004)

Tognini-Bonelli (2001) distingue ainda, os *corpora* multilíngues entre *corpora* de tradução, em que estão os *corpora* paralelos e *corpora* comparáveis. Os *corpora* paralelos são compostos por determinado texto e sua respectiva tradução ou traduções. Podem passar pelo processo de alinhamento, já que visam a busca pelos correspondentes tradutórios. Já os *corpora* comparáveis, são textos escolhidos pela similaridade, tipologia textual e que também permitem a identificação de correspondentes, porém não há o processo de alinhamento. (TOGNINI-BONELLI, 2001)

Partilhando da mesma ideia, Tagnin (2015) define os *corpora* paralelos, como compostos por textos originais em determinada língua e sua respectiva tradução. Geralmente utilizados para análise de obras literárias em que o pesquisador compara uma obra literária com sua tradução em determinada língua, ou compara o original com diversas traduções feitas sobre a mesma obra.

Já os *corpora* comparáveis, são compostos por textos originais nas duas (ou mais) línguas em estudo. Para que seja possível a comparação, os *corpora* devem ser constituídos de textos de mesma tipologia, temática e abranger período semelhante. (TAGNIN, 2015)

Ressaltamos que os *corpora* utilizados nesta pesquisa se classificam como *corpora* comparáveis, uma vez que o objetivo é realizar um estudo comparativo dos termos da área de Educação mais frequentes em artigos de dois jornais brasileiros e dois jornais norte-americanos.

A tipologia dos *corpora* utilizados em nossa pesquisa apresenta-se da seguinte maneira:

- a) **De língua escrita:** textos escritos em língua portuguesa e língua inglesa retirados dos jornais *O Estadão*, *O Globo*, *The New York Times* e *Los Angeles Times*.
- b) **Sincrônico e contemporâneo:** compreende um período de tempo curto (junho a agosto de 2017)
- c) **Equilibrado:** possui a mesma quantidade de textos: 200 textos escritos em língua portuguesa e 200 textos escritos em língua inglesa.
- d) **Bilíngue:** textos escritos nas línguas portuguesa e língua inglesa.
- e) **Comparáveis:** o *corpus* em língua portuguesa é comparável ao *corpus* de língua inglesa, dada a similaridade dos textos que formam os *corpora*.
- f) **Especializados:** abrange a área específica, que são artigos jornalísticos da área de educação.
- g) **De estudo:** *corpus* compilado para os objetivos de nossa pesquisa: levantamento do léxico representativo da área de educação em artigos jornalísticos.

Destacamos, assim como Pearson (1996 apud ZANETTIN, 1998) a relevância dos *corpora* especializados, que são uma fonte útil para a terminologia. Ainda segundo o autor, um *corpus* comparável especializado oferece uma gama de informações e expressões dentro de certos contextos (ZANETTIN, 1998, p. 6), com isso, justificamos a escolha e compilação dos *corpora* comparáveis para a realização da presente pesquisa e passamos a discussão acerca da ferramenta computacional utilizada neste estudo.

2.1.2. O Programa *WordSmith Tools* e suas Ferramentas

O site de apresentação do *WordSmith Tools*, o define como um programa de análise lexical, utilizado na exploração de *corpora*. Desenvolvido por Mike Scott (2012) e comercializado pela *Oxford University Press*, contribui para as produções lexicográficas da própria editora e de professores e pesquisadores há vinte e três anos, estando atualmente em sua sétima versão. Nesta pesquisa, ao referirmos ao *WordSmith Tools*, tratamos da versão 6.0, utilizada para a obtenção dos dados.

De modo geral, este tipo de programa permite-nos a obtenção de dados baseados na frequência e coocorrência de palavras em *corpora*, por meio de ferramentas que executam tarefas como a produção de listas de palavras, listas de palavras-chave e concordâncias por meio de cálculos estatísticos. Importante salientar que o computador ou a ferramenta computacional não substitui o papel do pesquisador, eles nos fornecem dados, mas o tratamento e análises continuam a ser nossas tarefas.

Os resultados obtidos através da utilização das ferramentas computacionais provem de cálculos estatísticos comparativos, que propiciam que os resultados sejam organizados de maneira rápida e eficiente.

O *WordSmith Tools* funciona com base em três princípios abstratos básicos, segundo Berber Sardinha (2004, p. 90):

- 1) Ocorrência: Observa-se aquilo que está presente no corpus. “Os itens devem estar presentes; itens que não ocorreram não são incorporados porque não são observáveis.”
- 2) Recorrência: Para que generalizações sejam possíveis é necessário que um item ocorra com certa frequência. “Os itens devem estar presentes pelo menos duas vezes.” Porém não se deve ignorar itens que apareçam apenas 1 vez. “Itens de frequência 1 são, em geral, raros (...)” e tem sua relevância.
- 3) Coocorrência: a frequência com que uma palavra aparece seguida de outra. A pesquisa determinará quantas palavras observar, se um conjunto de duas, três ou mais palavras. “Os itens devem estar na presença de outros. Um item isolado é muito pouco informativo”.

Estes princípios mantêm estreita relação com as ferramentas disponíveis no programa, visualizadas na Figura 3, que apresenta a janela inicial do *WordSmith Tools*,

versão 6.0. As três ferramentas principais (*Concord*, *KeyWords* e *WordList*) serão explicadas a seguir:

Figura 3 - Tela inicial do WordSmith Tools, versão 6.0



Fonte: PrintScreen do WordSmith Tool 6.0

- *Wordlist* (Lista de palavras)

De acordo com Berber Sardinha (2004, p. 91), A ferramenta *Wordlist* (lista de palavras), propicia a criação de listas de palavras a partir do corpus de estudo. Para melhor esclarecer esta funcionalidade, daremos como referência esta pesquisa, composta de quatro *corpora* comparáveis, cada um contendo artigos jornalísticos da área de educação. Sendo dois *corpora* comparáveis de língua portuguesa, cada um contendo cem artigos jornalísticos da área de educação de dois jornais diferentes. Da mesma forma, foram utilizados dois *corpora* comparáveis de língua inglesa, compostos por artigos originalmente escritos em língua inglesa na área de educação. Portanto, quatro listas de palavras foram criadas. De cada *corpus*, o programa produziu duas listas de cada vez, uma ordenada alfabeticamente e outra por ordem de frequência das palavras. Além disso, o programa gerou uma terceira janela composta por dados estatísticos a respeito dos dados. As listas de palavras foram utilizadas para a criação das listas de palavras-chave, que são objetos de nossa pesquisa.

A lista de palavras disposta em ordem alfabética e frequencial apresenta os elementos:

- **Word:** as palavras contidas nos textos
- **Freq.:** a frequência/quantidade que cada item ocorreu no texto
- **%:** a porcentagem do total de itens do texto a que corresponde cada item
- **Texts:** a quantidade de textos que apresenta o item
- **Lemmas:** outros itens cujas frequências foram adicionadas ao item corrente. São itens que incorporam formas derivadas. Por exemplo: o lema estudar pode compreender as formas estudo, estudei, estudando, estudado. Há a opção de lematizar ou não, ficando a critério do pesquisador. No caso desta pesquisa, não utilizamos a lematização.

Na Figura 4 vemos as listas de frequência dos corpora comparáveis em português. A primeira tabela trata-se da lista do jornal *O Globo* e a segunda tabela do jornal *O Estadão*. Observa-se que as primeiras palavras que compõem as listas são gramaticais em ambos os jornais. Na primeira posição aparece a preposição *de* (2950 + 2850 ocorrências). A primeira palavra de conteúdo aparece na 23ª posição no jornal *O Globo* e na 20ª posição no *Estadão*. Em ambos os jornais a primeira palavra de conteúdo é **educação**, tema de nossa investigação.

Figura 3 - Lista de palavras dos jornais *O Globo* e *O Estadão*

Word	Freq.	Word	Freq.		
1	DE	2950	1	DE	2850
2	A	1953	2	A	2076
3	O	1897	3	E	1748
4	QUE	1738	4	O	1575
5	E	1535	5	#	1486
6	DO	1184	6	QUE	1349
7	DA	941	7	DO	1051
8	#	933	8	PARA	908
9	PARA	930	9	EM	804
10	EM	804	10	DA	785
11	É	681	11	COM	638
12	NO	642	12	É	632
13	COM	616	13	OS	548
14	OS	612	14	NO	504
15	UMA	540	15	NA	471
16	UM	522	16	UMA	470
17	NÃO	515	17	AS	461
18	NA	444	18	UM	445
19	AS	391	19	NÃO	403
20	SE	372	20	EDUCAÇÃO	388
21	POR	336	21	MAIS	338
22	MAIS	323	22	SE	333
23	EDUCAÇÃO	320	23	DOS	306
24	DOS	311	24	POR	296
25	COMO	276	25	ENSINO	280
26	AO	257	26	ALUNOS	272
27	ALUNOS	231	27	COMO	269
28	MAS	221	28	DAS	248
29	DAS	197	29	SÃO	232
30	SÃO	196	30	ESCOLA	212
31	ESCOLA	194	31	ANOS	181

Fonte: PrintScreen do *WordSmith Tool 6.0*

Em língua inglesa, conforme Figura 4, repetem-se as palavras gramaticais ocupando as primeiras posições. Em primeiro lugar aparece o artigo definido *the* (4.204

+ 6.084 ocorrências) nos jornais *Los Angeles Times* e *The New York Times*, respectivamente. A primeira palavra de conteúdo no jornal *Los Angeles* é *students* (13º posição) seguida de *school* (19º posição), em contrapartida, no jornal *The New York Times* inverte-se a ordem, *school* aparece na 19º posição, seguida de *students* na posição seguinte.

Notamos que há semelhança entre os jornais da mesma língua, visto que ambos os artigos escritos em língua portuguesa a primeira palavra de conteúdo é *educação*, diferente dos jornais norte-americanos que apesar de combinarem entre si, resultando nas palavras *school* e *students*, diferem do resultado disposto em língua portuguesa.

Figura 4 - Listas de palavras dos jornais *Los Angeles Times* e *New York Times*

Word	Freq.	Word	Freq.		
1	THE	4204	1	THE	6084
2	TO	2327	2	TO	3182
3	A	2146	3	A	3164
4	AND	2067	4	OF	3014
5	OF	1945	5	AND	2825
6	IN	1607	6	IN	2334
7	#	1235	7	THAT	1536
8	THAT	1008	8	#	1364
9	FOR	894	9	FOR	1171
10	S	772	10	S	1144
11	SAID	769	11	IS	902
12	ON	606	12	ON	887
13	STUDENTS	586	13	SAID	883
14	WAS	537	14	IT	862
15	WITH	534	15	WAS	844
16	IS	532	16	I	754
17	HE	511	17	AT	753
18	AT	483	18	WITH	708
19	SCHOOL	483	19	SCHOOL	677
20	AS	470	20	STUDENTS	646
21	IT	438	21	AS	625
22	SHE	378	22	HE	548
23	THEY	362	23	ARE	538
24	AN	349	24	BUT	512
25	ARE	347	25	BY	504
26	HAVE	346	26	BE	488
27	BY	343	27	HAVE	484
28	WHO	339	28	SHE	481
29	FROM	337	29	FROM	475
30	BE	315	30	WHO	474
31	WERE	300	31	WERE	470

Fonte: PrintScreen do *WordSmith Tool 6.0*

Há ainda os dados estatísticos gerados pelo *WordSmith Tools*, que são apresentados em uma janela com os seguintes títulos:

- **Text File:** nome do arquivo

- **Tokens:** número de itens ou ocorrências no corpus. Por exemplo, a frase *professora do ensino médio e ensino fundamental*, possui 7 itens: professora (1), do (2), ensino (3), médio (4), e (5), ensino (6), fundamental (7)

- **Types:** número de formas ou vocábulos, ou seja, itens sem repetição. A frase *professora do ensino médio e ensino fundamental*, apresenta 6 itens. Há a ocorrência da palavra *ensino* duas vezes.

- **Type-Token Ratio:** a razão forma/item, resultado do cálculo em que se divide o total de formas pelo total de itens, transformando o valor em porcentagem. Esse resultado indica a riqueza lexical do texto, pois quanto maior o seu valor, mais palavras diferentes o texto conterà. Já um valor baixo indicará um número alto de repetições e consequentemente pode indicar um texto menos variado em relação ao vocabulário utilizado.

- *KeyWords* (Palavras-chave)

A ferramenta *KeyWord* (palavras-chave), contrasta uma lista de palavras de um *corpus* de estudo com uma lista de palavras de um *corpus* de referência. As palavras consideradas com maior chavicidade (ou palavras-chave), conforme Berber Sardinha (2004), são aquelas que possuem representatividade no corpus de estudo, o que não está relacionado apenas com a alta frequência da palavra. Para alcançar o resultado das palavras que sejam realmente representativas, é preciso comparar a lista de palavras mais frequentes no corpus de estudo com a lista de palavras mais frequentes de um corpus de referência. Para isso, é necessário que:

1. Um corpus de estudo [principal], representado por 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.
2. 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 análise. A sua função é fornecer uma norma com a qual se fará a comparação das frequências do corpus de estudo. A comparação é feita por meio de uma prova estatística selecionada pelo usuário (qui-quadrado ou log-likelihood). As palavras cujas frequências 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)

Dessa forma, para o levantamento das palavras-chave em português, foi utilizado o *corpus* de referência *Lácio-Ref*, que possui 4.156.816 itens, composto por textos organizados em cinco **gêneros** (informativo, científico, prosa, poesia e drama),

diversos **tipos de textos** (reportagens, artigos, crônicas, cartas), vários **domínios** (como educação, engenharia, política) e alguns **meios de distribuição** (revista, internet, livro).¹⁴

A lista de palavras mais frequentes do nosso *corpus* de estudo foi comparada com uma lista de palavras mais frequentes do *corpus* de referência, na qual o programa *WordSmith Tools* realiza um cálculo estatístico, resultando na chaticidade do termo. De modo geral, quando a palavra aparece mais vezes no *corpus* de estudo do que no *corpus* de referência, é considerada uma palavra-chave. Este resultado é dado de forma automática por meio do programa de análise lexical *WordSmith Tools*.

Para o levantamento das palavras-chave em língua inglesa, utilizamos como *corpus* de referência, o *BNC (British National Corpus)* que contém 100 milhões de palavras de variadas fontes. Sua última edição foi lançada em 2007. O *corpus* inclui textos de jornais, periódicos, livros acadêmicos, ficção, cartas, memorandos, artigos, entre outros tipos de textos.¹⁵

Para o desenvolvimento desta pesquisa, interessa-nos as listas de palavras-chave, produzidas a partir dos dois jornais brasileiros (*O Globo* e *O Estadão*) e dos dois jornais norte-americanos (*The New York Times* e *Los Angeles Times*).

- *Concord* (Lista de concordância)

A ferramenta *Concord* produz listas de concordância ou listas de ocorrências de um item específico, acompanhado de texto antes e depois. Com esta ferramenta é possível verificar aspectos específicos da linguagem, como o uso de colocações fixas, entre outras questões visíveis dentro de um pequeno contexto. Nesta pesquisa ocupamo-nos do tratamento de termos simples e termos complexos, portanto as listas de concordância auxiliam-nos na identificação de possíveis termos complexos. Os termos analisados são da classe dos substantivos, partimos do substantivo simples e no momento da análise verifica-se se é um termo complexo e qual sua correspondência na língua estrangeira.

A Figura 6 apresenta as linhas de concordância a partir da busca por **ensino**, que se encontra entre as primeiras palavras da lista de palavras-chave, tanto no jornal *O Globo* quanto no jornal *O Estadão*. No *corpus* do jornal *O Globo*, nota-se que **ensino** ora é termo simples, ora termo complexo. Repetidas vezes a palavra **ensino** coocorre com as palavras

¹⁴ Informações disponíveis em: <<http://143.107.183.175:22180/lacioweb/plancamento.htm>> Acesso em: 10 de janeiro de 2019.

¹⁵ Informações disponíveis em: <<http://www.natcorp.ox.ac.uk/corpus>> Acesso em: 10 de janeiro de 2019.

fundamental, médio, superior, a distância, que são modalidades do sistema educacional brasileiro. Nesta pesquisa, a ferramenta *Concord* nos auxiliou na identificação e nas diferentes possibilidades combinatórias dos termos.

Figura 5 - Linhas de concordância a partir da busca por 'ensino'

Concordance

a Helena defende que parte dos investimentos do MEC no ensino superior seja repassada ao Ministério da Saúde, a exemplo dos cerca de 1,5 bilhões de reais, um indicador que mede a qualidade das instituições de ensino superior brasileiras, a Uerj registrou nota 4 em uma escala na qual a nota máxima é 5. No ano de 2015, o ensino médio teve uma taxa de evasão de 16,8%. Levando em consideração todas as séries do ensino médio, a evasão corresponde a 11% dos alunos do ciclo. Em relação ao ensino fundamental, uma turma do 2º ano do ensino fundamental ficou dividida. Pouco mais da metade dos pais [16 em relação ao número maior de mudança para a modalidade no final do ensino fundamental e que essa transição é maior na rede municipal, nos anos iniciais do texto da Base Nacional Comum Curricular (BNCC) do ensino fundamental os termos "identidade de gênero" e "orientação sexual" arde nas — lembra Junior, que largou os estudos no 2º ano do ensino médio para trabalhar e hoje, aos 31 anos, cursa o 3º ano e pretende ser professor pedagogo que trabalham o tema nos anos iniciais do ensino fundamental receberão material. Haverá livros também para os professores e para os alunos. O fato de o candidato ter optado pelo ensino a distância não faz diferença. O que realmente importa está em três pontos. Em relação ao período de 2014/2015 e nos anos iniciais do ensino fundamental, o Piauí, a Bahia e o Pará são os estados com o maior percentual de abandono para 2016, o percentual de alunos que abandonaram o ensino médio na rede pública praticamente dobrou: foi de 4,5% para 8,5%. O índice em 2008. Naquele ano, 17% dos alunos haviam abandonado o ensino médio na rede pública no estado, percentual que superava a média nacional [12,2%], 7º ano [11,7%], 8º ano [9%] e 9º ano [8,2%] do ensino fundamental. Já a evasão é mais expressiva nos anos finais do ensino fundamental. Já a evasão é mais expressiva nos anos finais do ensino fundamental - 9º ano [7,7%], 8º ano [4,9%], 7º e 6º ano [4,7%] - e no último ano [6,8%]. Na segunda posição de reprovação está o 6º ano do ensino fundamental, com 14,4%. Já o segundo lugar para a evasão vai para o 2º ano do ensino fundamental, sobretudo na passagem dos anos iniciais para os finais, e não para o segundo segmento, que tem sido esquecido. Fala-se do ensino médio, da alfabetização, mas os anos finais do ensino fundamental têm sido negligenciados. Já o segundo lugar para a evasão vai para o 2º ano do ensino médio, com 12,7%. Alto percentual de reprovação também é visto no 3º ano do ensino médio. No Caderno Digital é oferecido apenas aos alunos do ensino médio, mas a partir do próximo ano também estará disponível para o ensino fundamental. Os dados Anísio Teixeira (INEP) revelam que o primeiro ano do ensino médio é a série escolar com o maior percentual de reprovação e evasão. O primeiro ano do ensino médio é a série com maior índice de reprovação e evasão. Inep divulgou dados inéditos sobre fluxo escolar; 6º ano do ensino fundamental tem a segunda maior taxa de reprovação BRASÍLIA. Dados do Inep sobre a educação single sex a partir do 2º ano do ensino fundamental até o ensino médio. Os alunos são divididos em duas unidades de ensino. — Esse empréstimo virá para apoiar a reforma do ensino médio, tendo como eixo a formação de professores das redes, a constituição de referência em todo o estado por seu programa de ensino da língua inglesa, o Bilingue Forever. Assim como Artur, os alunos do ensino médio que atendem alunos mais carentes também costumam ser mais raros. Segundo o Enem, menos 10% dos inscritos para certificação do ensino médio eram aprovados, isso é um dado importantíssimo que nos obriga a pensar em melhorias. Aline Pereira, gerente de produtos do Stoodi, plataforma de ensino à distância com foco no Enem e pré-vestibular, diz que o número de alunos do ensino médio, da alfabetização, mas os anos finais do ensino fundamental têm sido um patinho feio, abandonado pela educação brasileira. O número de alunos de 114.825 escolas. As escolas privadas do ensino médio terão de terça-feira [27] até 14 de julho para assinar o termo de adesão, mas de forma equilibrada, defende Vicente. O diretor de ensino médio dá uma dica especial para quem já terminou o ensino médio e estudou em uma escola pública. Do total dessas reinseridas no sistema de ensino, apenas 10% novamente abandonaram a escola. Leituras simplificadas do p

Fonte: *PrintScreen do WordSmith Tool 6.0*

Fica clara, a necessidade do uso de tais ferramentas para o desenvolvimento de uma pesquisa deste tipo, capaz de nos fornecer dados, por vezes imperceptíveis aos olhos humanos. Tendo definido as principais funcionalidades do programa *WordSmith Tools*, bem como a funcionalidade de suas ferramentas, passamos a conceituação das ciências do léxico, áreas que associadas a LC dão sustentação teórica a esta pesquisa e corroboram para o alcance dos objetivos propostos.

2.2. Lexicologia, Terminologia, Lexicografia e Terminografia

Nesta pesquisa temos como um dos objetivos a identificação de termos da área da educação, sendo assim, apresentamos a conceituação e a distinção entre palavra, vocábulo e termo. Posteriormente, trataremos das ciências do léxico, ou seja, da Lexicologia, Terminologia, Lexicografia e Terminografia.

O Léxico de uma língua é compreendido como um sistema de signos utilizado pelos falantes, independentemente de sua cultura ou heterogeneidade, assim como esclarece Biderman (1978, p. 139):

O léxico de qualquer língua constitui um vasto universo de limites imprecisos e indefinidos [que] abrange todo o universo conceptual dessa língua. Qualquer sistema léxico é a somatória de toda a experiência acumulada de uma sociedade e do acervo da sua cultura através das idades. (BIDERMAN, 1978, p. 139).

É através desse sistema de signos que nos comunicamos, nomeamos, classificamos:

[...] o léxico de uma língua constitui uma forma de registrar o conhecimento do universo. Ao dar nomes aos referentes, o homem os classifica simultaneamente. Assim, a nomeação da realidade pode ser considerada como a etapa primeira no percurso científico do espírito humano de conhecimento do universo. (BIDERMAN, 1998, p. 91)

Ou seja, o léxico é composto pelas palavras e “a palavra é a pedra de toque¹⁶ da linguagem humana”. (BIDERMAN, 1998). Como há divergências, principalmente, entre os sentidos atribuídos por ‘palavra’, ‘vocábulo’ e ‘termo’, faz-se necessária a conceituação e a distinção para instruir esse trabalho.

Para Barros (2004, p. 40) “a “palavra” é uma unidade léxica (ou unidade lexical), ou seja, é um signo linguístico, composto de expressão e de conteúdo, que pertence a uma das grandes classes gramaticais”, que no caso da língua portuguesa são dez. Nesta pesquisa nos limitaremos à classe dos substantivos.

Para elucidar o trabalho, trazemos as dez classes gramaticais e respectivas explicações: **1) Substantivos:** palavras que dão nome aos seres em geral, sejam eles animados ou inanimados. Os substantivos podem ser classificados como Substantivo comum; substantivo próprio; substantivo concreto; Substantivo abstrato; Substantivo

¹⁶ Segundo o Dicionário Aurélio (FERREIRA, 2010), a pedra de toque é empregada pelos joalheiros para avaliar a pureza dos metais. No sentido figurado, significa o meio de avaliar e aferir.

simples; Substantivo composto; Substantivo primitivo; substantivo derivado; Substantivo coletivo. **2) Verbo:** consiste na palavra variável que exprime um processo que se passa no tempo, indicando: Ação; Estado; Mudança de estado; Fenômeno da natureza; Existência; Desejo; ou Conveniência: convém. **3) Adjetivo:** palavra utilizada para caracterizar o substantivo ou qualquer palavra que possui valor de substantivo, indicando-lhe um atributo, estado, propriedade, aspecto ou maneira de ser. **4 – Pronome:** consiste numa palavra variável em gênero, número e pessoa que acompanha ou representa o substantivo, indicando-o como pessoa no discurso ou situando-o no espaço e no tempo. Os pronomes são classificados em: Pronomes pessoais: eu, tu, ele, ela, nós etc.; Pronomes de tratamento: Vossa Alteza, Vossa Senhoria, Vossa Excelência etc. **5 – Artigo:** consiste na palavra que se antepõe a um substantivo, determinando-o de maneira precisa conforme suas flexões, indicando gênero e número. Os artigos podem ser classificados em Artigos definidos: o, a, os, as; Artigos indefinidos: um, uma, uns, umas. **6 – Numeral:** Trata-se da palavra utilizada para indicar a quantidade exata de seres ou a posição que um ser ocupa numa determinada série. Podem ser cardinais, ordinais, multiplicativos ou fracionários. **7 – Preposição:** é a palavra invariável responsável por ligar dois termos da oração, subordinando um ao outro. **8 – Conjunção:** consiste na palavra invariável que tem a função de ligar palavras, grupos de palavras, orações e frases, exprimindo uma relação de sentido entre as unidades ligadas. **9 – Interjeição:** palavra invariável que serve para exprimir emoções e sentimentos súbitos. **10 – Advérbio:** palavra invariável que serve para modificar o verbo, exprimindo determinada circunstância (lugar, tempo causa etc.). Podem ser de afirmação, dúvida, intensidade, lugar, tempo, modo ou negação.

Destacamos que a análise ora realizada, limita-se a um *corpus* formado de palavras substantivas oriundas dos artigos jornalísticos da área de Educação provenientes de quatro jornais (*O Estadão, O Globo, The New York Times e Los Angeles Times*)

Na LC e análise estatística, “palavra” é o termo que denomina a quantidade de palavras-ocorrência em um texto. A “palavra” é considerada e contabilizada na análise estatística cada vez que há um espaço em branco, ou seja, conta-se cada bloco de letras que aparece entre espaços. Essa é a principal característica que diferencia “palavra” e “vocábulo”, já que este último se refere, para efeito de análise de corpus linguístico, às palavras sem repetição (BARROS, 2004). Nosso *corpus* de estudo contém 118.606 palavras e 15.949 vocábulos.

Para exemplificar essa distinção, vejamos o trecho a seguir:

O atual governo tem como promessa a criação de 65 mil **vagas** em creche (que atende crianças de 0 a 3 anos) até março de 2018. Por enquanto, nenhuma nova creche foi inaugurada, mas a Secretaria de Educação diz que foram abertas 4,5 mil novas **vagas**, com o uso de **vagas** ociosas nas unidades.

Neste trecho, o programa contabilizará **vagas** três vezes e **creche** duas vezes, como ‘palavras’, no total de ocorrências, ou seja, mesmo que se repitam, serão contadas como uma nova ocorrência, como novo item. Como ‘palavras’, o trecho destacado contabilizará 54 *tokens* (itens), como ‘vocábulos’, o trecho contém 51 *types* (formas), pois **vagas** e **creche** serão contados apenas uma vez.

A relação “vocábulo” e “termo”, dá-se como nos ensina Barros (2004):

[...] termo é um vocábulo, uma vez que é um modelo de realização lexical no texto. Seu caráter de termo se dá pelo fato de que designa um conceito específico de um domínio de especialidade. O conjunto terminológico presente nesse texto constitui, na verdade, um subconjunto do conjunto vocabular do mesmo. Assim, um termo é também um vocábulo, além de ser uma palavra. (BARROS, 2004, p.42)

O termo pode ser classificado de acordo com sua formação (formado por letras ou formas braquigráficas), conforme detalharemos a seguir. Quando formado por letras, recebe ainda classificação de acordo com a formação, assumindo a designação de: termo simples, termo complexo ou termo composto, assim como esclarece Barros (2004):

a) **Formado apenas por letras:** “constituído de um só radical com ou sem afixo” (ISO 1087, 1990, p. 7). O termo é composto por um único lexema, chamado de termo simples. Como exemplo podemos citar: “educação”, “alunos”, “escolas”. Os termos compostos são formados por dois ou mais radicais e se distinguem dos termos complexos devido ao alto grau de lexicalização. Em situação de não autonomia são grafados com hífen, como ocorre em “mão-de-obra”. (BARROS, 2004, P. 100)

Os termos complexos são “constituídos de dois ou mais radicais, aos quais podem-se acrescentar outros elementos” (ISO 1087, 1990, p. 7 APUD BARROS, 2004, p. 100), como exemplo há “progressão continuada”, “carga horária”, “educação básica”.

É comum ao discurso especializado que ocorra um processo de criação de termos com base em oposições distintas que tem o mesmo lexema base, chamado de composição sintagmática, como podemos exemplificar com o conjunto de sintagmas terminológicos: (BARROS, 2004, p. 101)

- ensino

- ensino médio
- ensino fundamental
- ensino infantil
- ensino superior
- ensino básico

Nessas composições cada termo atribui um significado distinto, tendo sempre como base para a composição, o termo “ensino”.

b) **Formas braquigráficas:** do grego “breve, curto” englobam as siglas e acrônimos. O acrônimo possui sua formação a partir das letras iniciais de um termo complexo e a pronúncia é de uma nova palavra. Exemplo: “MEC” (Ministério da Educação); “ENEM” (Exame Nacional do Ensino Médio); “SiSU” (Sistema de Seleção Unificada). Já a sigla é pronunciada letra por letra. Exemplos “PNE” (Plano Nacional de Educação); “PNLD” (Programa Nacional do Livro Didático).

A partir de tais considerações, passamos ao estudo das Ciências do Léxico, Lexicologia, Terminologia, Lexicografia e Terminografia.

A ciência responsável pelo estudo científico do léxico é a Lexicologia. A unidade padrão da Lexicologia é a unidade lexical, caracterizada pela não-separabilidade dos elementos que a realizam. Pode ser estudada por meio do eixo paradigmático (eixo das substituições) e por meio do eixo sintagmático (eixo das combinações). (BARROS, 2004, p. 60)

Do ponto de vista descritivo, a Lexicologia ocupa-se de variadas tarefas, segundo Barbosa (1991, p. 183)

1. Definir conjuntos e subconjuntos lexicais, como o universo léxico, conjunto vocabulário, léxico efetivo e virtual, vocabulário ativo e passivo.
2. Conceituar e delimitar a unidade lexical de base (a lexia), bem como elaborar os modelos teóricos subjacentes às suas diferentes denominações;
3. Analisar e descrever as estruturas morfo-sintático-semânticas de tais unidades, sua estruturação, tipologia e possibilidades combinatórias
4. Examinar essas unidades de acordo com sua carga ideológica, força persuasiva, e natureza modelizante;
5. Abordar a palavra como um instrumento de construção e detecção de uma “visão de mundo”, de uma ideologia, de um sistema de valores, como geradora e reflexo de recortes culturais;
6. Analisar a influência do contexto em cada palavra;
7. Analisar e descrever as relações entre a expressão e o conteúdo das palavras e os fenômenos decorrentes, como: polissemia, homonímia, homossemia, parassinonímia, hiperonímia.
8. Estudar o conjunto de palavras de determinado sistema.

Já no ponto de vista aplicado, Barbosa (1991) situa a contribuição da Lexicologia nos campos do ensino de línguas, seja materna ou estrangeira, processos de tradução automática ou não, entre outras.

Tais tarefas, sejam elas no campo da Lexicologia descritiva quanto da Lexicologia prática (aplicada), podem valer-se do aparato teórico-metodológico da LC, capaz de extrair dados para que os lexicólogos executem suas análises, sejam elas voltadas para a definição de conjuntos vocabulários, para a aplicação no processo de ensino-aprendizagem ou ainda para a tradução.

Como destacado, a Lexicologia lida com o léxico geral da língua, preocupando-se em descrever a língua como um todo em determinado contexto. Já a Terminologia lida com o léxico especializado. O conjunto de termos de determinada área ou subárea, designa-se terminologia (grafaremos com minúscula). Já o estudo de tais terminologias, bem como métodos e pressupostos que permitem a descrição dessa linguagem, designamos de Terminologia (grafada com maiúscula), por sua natureza científica. Assim, nosso trabalho atém-se à terminologia da área de educação, a partir de artigos jornalísticos.

Pavel e Nolet (2001, p.17), assinalam que ‘terminologia’ significa um “conjunto de palavras técnicas pertencentes a uma ciência, uma arte, um autor ou grupo social”. Os estudos em Terminologia começaram com a TGT e desenvolveram-se para a TCT, conforme esclarecemos a seguir.

Barros (2004) afirma que as origens da Terminologia se encontram no Ocidente, nos trabalhos de Eugen Wuster, que na década de 1930 estabeleceu pressupostos da chamada Teoria Geral da Terminologia (doravante TGT). Em torno disso desenvolveu-se a Escola de Viena, de Praga e a Russa que visavam à padronização terminológica em busca da perfeita comunicação científica em todo o mundo.

Para o engenheiro austríaco, a terminologia tinha como objetivo dar as bases científicas para a eliminação de qualquer ambiguidade dos discursos teóricos e científicos. Em sua visão, o termo era dissociado, separando o léxico da gramática, contexto e discurso, como se fossem unidades com vida independente. Nesta perspectiva não deveria existir termo polissêmico, sinônimo ou homônimo. (BARROS, 2004, p. 55)

Ainda segundo a autora, a Terminologia pode ainda ser classificada como Terminologia descritiva, em que se ocupa da coleta e da descrição dos termos por meio das definições, geralmente resultando na produção de vocabulários ou glossários; e a

Terminologia Normativa, que se ocupa da uniformização de termos ou da criação e atribuição de novos termos. (BARROS, 2004)

Na perspectiva de Wuster, o termo, visto de uma perspectiva terminológica normativa, seguia o pressuposto de denominações de conceitos, em que cada conceito corresponderia a uma única denominação. Não sendo considerados como elementos naturais da língua, mas sim, identificados através de rótulos ou etiquetas. (KRIEGER & FINATTO, 2004)

A TGT perdurou por pelo menos 70 anos, os princípios e os métodos terminológicos serviram e servem de base a todo trabalho terminológico. Porém, com o crescimento das produções na área, principalmente ao que concerne à descrição de conjuntos terminológicos houve um processo de reflexão sobre o fazer terminológico e as limitações da TGT.

A necessidade de uma nova proposta terminológica conduz à proposta de Cabré (1999), chamada de Teoria Comunicativa da Terminologia (doravante TCT). O principal questionamento da autora recai sobre a independência do termo, ignorando a cultura e outros aspectos da língua.

Na TCT, não há distinção drástica entre a unidade terminológica (termo) da unidade lexical da língua geral (palavra), visto que os termos são unidades linguísticas que não deixam de ser signos de uma língua geral. (BARROS, 2004).

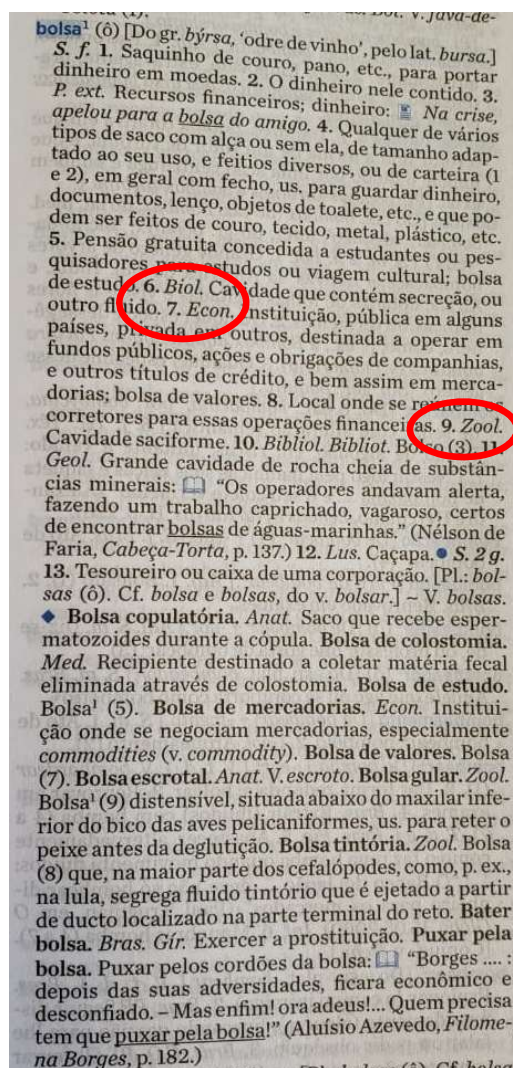
São unidades léxicas (como todo signo linguístico) de um dado código e seu caráter de termo é ativado de acordo com seu uso em um contexto e situação adequados. Essa ativação consiste em uma seleção dos módulos de traços apropriados, que incluem os traços morfossintáticos gerais da unidade e uma série de traços semânticos e pragmáticos específicos que descrevem seu caráter de termo dentro de um determinado domínio. (CABRÉ, 1999, p. 123 APUD BARROS, p. 57)

Dessa forma, contrária ao pressuposto da TGT em que a forma e o conteúdo são independentes, a TCT considera o signo terminológico como uma unidade composta de forma e conteúdo. Para ela fora do contexto, as unidades léxicas são apenas unidades léxicas, assumindo o valor de termo quando utilizada em um contexto específico. (CABRÉ, 1999, p. 123 APUD BARROS, p. 57).

Essa questão fica clara ao lidarmos com nosso corpus de estudo, que recorta um contexto específico: a educação. Observa-se por exemplo, a partir das palavras-chave dos *corpora* em língua portuguesa sobre educação, o termo **bolsa**, que apresenta uma frequência de 62 ocorrências.

Ao realizar a consulta no dicionário Aurélio (FERREIRA, 2010), a palavra bolsa, na língua geral, perpassa pelas definições de: 1) saquinho de couro ou pano, para portar dinheiro ou moedas; 2) pensão gratuita para estudantes; 3) Instituição destinada a operar fundos públicos; 4) local onde se reúnem os corretores; enfim, há uma série de acepções para uma mesma palavra. Sendo assim, podemos identificar que a acepção 1 é usada na língua geral, porém a acepção 2 relata um significado pertinente apenas ao ambiente educacional, ou seja, torna-se um termo devido à sua especialidade. Da mesma forma a acepção 3 e 4 definem bolsa na especialidade econômica, muitas vezes, o termo é identificado no dicionário com a sigla correspondente à sua especialidade, como observa-se na figura 7:

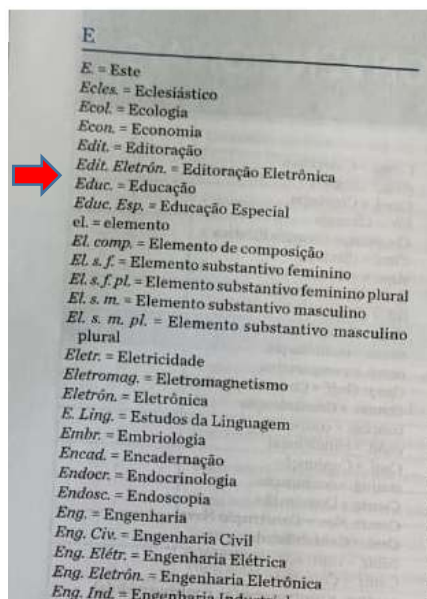
Figura 6 - Verbetes 'bolsa' no Dicionário Aurélio com destaque às siglas de especialidade



Fonte: Dicionário Aurélio (FERREIRA, 2010)

Embora o Dicionário Aurélio (FERREIRA, 2010) traga algumas acepções com a respectiva sigla referente à especialidade, a acepção 5, que trata no termo no âmbito da educação, não possui qualquer indicação à área. A parte inicial do Dicionário apresenta as siglas que são utilizadas na obra e a sigla Educ. (conforme Figura 8) apresenta-se entre as possibilidades, porém não foi utilizada neste caso.

Figura 7 - Siglas no Dicionário Aurélio e destaque para a sigla de 'educação'



Fonte: Dicionário Aurélio (FERREIRA, 2010)

Sendo assim, partilhamos dos princípios da TCT e dos pressupostos teóricos e metodológicos apresentados por Cabré (1999) para a coleta e análise dos dados, pois na TCT, o termo passou a ser visto dentro de um texto, possuindo não só uma dimensão linguística, mas também pragmática, discursiva e comunicativa. (BARROS, 2007)

A TCT articula-se baseada na valorização dos aspectos comunicativos das linguagens especializadas em detrimento dos propósitos normalizadores, bem como na compreensão de que as unidades terminológicas formam parte da linguagem natural e da gramática das línguas. De acordo com o princípio comunicativo, uma unidade lexical pode assumir o caráter de termo em função de seu uso em um contexto e situação determinados. (KRIEGER & FINATTO, 2004, p. 35)

Desta forma, busca-se por meio da Terminologia, a padronização do léxico especializado, facilitando a comunicação internacional. Os terminólogos produzem glossários e dicionários técnicos que visam auxiliar o tradutor, os aprendizes, professores de idiomas o público em geral, tendo em vista a árdua tarefa da sistematização do léxico

tomado como “termos que transmitem conteúdos próprios de cada área. ” (KRIEGER; FINATTO, 2004, p. 17)

Também associada a evolução tecnológica e a proliferação de textos científicos e técnicos, a Terminologia possui não só o propósito de denominar o conteúdo especializado, mas também disseminar esse conhecimento.

Como discutido, tanto a Lexicologia quanto a Terminologia tratam da descrição da linguagem. Aquela, abordando questões da língua geral e esta, lidando com aspectos das línguas de especialidade.

Já a produção de materiais, é tarefa da lexicografia e da terminografia. Da mesma maneira, a lexicografia lida com a língua geral, portanto seus produtos visam a elaboração de dicionários de língua ou especiais, registram as unidades lexicais em todas as suas variações e acepções (a citar o exemplo de *bolsa*, em que o dicionário Aurélio, por ser um dicionário padrão da língua utiliza todas as acepções da palavra, incluindo as suas acepções terminológicas). Já os produtos terminológicos (vocabulários ou glossários) registram os termos, os modelos de realização lexical em nível das normas de universo de discurso especializado em um contexto específico (em nossa pesquisa terminológica, interessa-nos a acepção e definição do termo *bolsa* aplicado ao contexto de educação).

A fim de esclarecer o que compete a cada área, elaboramos o quadro a seguir adaptado de Barros (2004, p. 64)

Quadro 4 - Sistematização das áreas das Ciências do Léxico

	Lexicologia	Terminologia	Lexicografia	Terminografia
<i>Campo de atuação</i>	Língua geral	Línguas de especialidade	Língua geral	Línguas de especialidade
<i>Unidade padrão</i>	Unidade lexical, lexema	Unidade terminológica, termo	Unidade lexical, lexema	Unidade terminológica, termo
<i>Nível de atualização da unidade lexical</i>	Sistema	Normas de especialidade	Sistema	Normas de especialidade
<i>Tipos de obras</i>			Dicionários de língua, dicionários especiais, dicionários de expressões idiomáticas, sinônimos, etc.	Vocabulários técnicos, científicos, especializados e glossários

Fonte: Quadro elaborado pela autora a partir de Barros (2004).

Dissertamos nas seções e subseções acima sobre os pilares teóricos que deram sustentação a esta pesquisa. A seção seguinte destina-se a discussão dos procedimentos metodológicos adotados nesta investigação, desde a compilação dos *corpora*, a seleção das palavra-chave e as análises das correspondências.

3. METODOLOGIA

A metodologia adotada nesta pesquisa é de natureza quantitativa e qualitativa recorre ao arcabouço teórico-metodológico da Linguística de *Corpus* (BERBER SARDINHA, 2000, 2004) para a obtenção e análise dos dados; aos Princípios da Lexicologia (BIDERMAN, 1985, 2002; WELKER, 2004) e da Terminologia (BARBOSA, 2001; BARROS, 2004) subsidiando a análise dos dados, com intuito de estudar o léxico representativo da área de Educação em artigos jornalísticos, propor correspondências e elaborar uma proposta de glossário bilíngue português-inglês a ser hospedado em uma plataforma online.

Para tornar possível, o percurso para esta pesquisa iniciou-se com a compilação dos *corpora* comparáveis, compostos da maneira como se segue:

1. Um *corpus* comparável, composto por 100 artigos jornalísticos escritos originalmente em português (57.593 itens), retirados do jornal brasileiro *O Estadão*, da seção de Educação, em sua versão *online*, no período de junho a agosto de 2017.
2. Um *corpus* comparável, composto por 100 artigos jornalísticos escritos originalmente em português (61.013 itens), retirados do jornal brasileiro *O Globo*, da seção de Educação, em sua versão *online*, no período de junho a agosto de 2017.
3. Um *corpus* comparável, composto por 100 artigos jornalísticos escritos originalmente em inglês (114.817 itens), retirados do jornal norte-americano *The New York Times*, na seção de Educação, em sua versão *online*, no período de junho a agosto de 2017.
4. Um *corpus* comparável, composto por 100 artigos jornalísticos escritos originalmente em inglês (80.815 itens), retirados do jornal norte-americano *Los Angeles Times*, na seção de Educação, em sua versão *online*, no período de junho a agosto de 2017.

Elencamos a seguir as etapas realizadas durante a pesquisa que serão detalhadas neste capítulo e posteriores:

- 1) Compilação dos quatro *corpora* comparáveis;
- 2) Criação das listas de palavras de cada jornal;
- 3) Criação das listas de palavras-chave de cada jornal;
- 4) Criação das linhas de concordância de cada jornal;
- 5) Seleção dos substantivos comuns em língua portuguesa e língua inglesa
- 6) Seleção dos substantivos comuns que se repetiram nos dois jornais de língua portuguesa e língua inglesa
- 7) Busca por correspondente Português-Inglês no corpus de estudo para atestar a escolha;
- 8) Consulta em dois dicionários bilíngues escolares *Longman Dicionário Escolar* e *Michaelis Dicionário Escolar* para verificar quais as possibilidades de equivalência propostas;
- 9) Consulta ao dicionário monolíngue *Longman Dictionary of Contemporary English*, buscando confirmar a definição em língua inglesa e verificar a correspondência escolhida, bem como, a consulta de sinônimos.
- 10) Consulta ao *Concord*
- 11) Consulta ao *Linguee*
- 12) Consulta ao *Education USA*
- 13) Criação de uma proposta de glossário

A construção dos *corpora* iniciou-se pela coleta e armazenamento dos artigos. Todos os artigos foram retirados da versão *online* dos jornais, sendo cada um salvo individualmente em pastas intituladas com o nome dos jornais e no formato .doc do *Word*, para uma melhor visualização e fácil localização das informações caso fosse necessária a busca manual.

Após a organização dos artigos, os mesmos foram salvos no formato “.txt” (texto sem formatação), ainda em quatro pastas individuais, cada qual contendo os cem artigos do respectivo jornal. No formato .txt, os artigos foram processados pelo programa de análise lexical *WordSmith Tools*, versão 6.0, de Scott (2012) definido como “um conjunto de programas integrados (‘suíte’) destinado à análise linguística ” (BERBER SARDINHA, 2009, p. 8) e utilizado na exploração de *corpora*. O programa possui três

ferramentas básicas utilizadas nesta pesquisa: *WordList*, em que se cria listas de palavras do *corpus* de estudo. O programa produz duas listas de cada vez, uma ordenada alfabeticamente e outra por ordem de frequência das palavras. A lista de frequência é utilizada para a criação da lista de palavras-chave; A *KeyWord*, contrasta a lista de palavras mais frequentes do *corpus* de estudo com a lista de palavras mais frequentes do *corpus* de referência. Resultando em uma lista de palavras-chave da área em estudo; e o *Concord* produz listas de concordância ou listas de ocorrências de um item específico, acompanhado de texto antes e depois, possibilitando verificar aspectos específicos da linguagem, como o uso de colocações fixas, entre outras questões visíveis dentro de um pequeno contexto. Em nossa pesquisa, as ferramentas foram utilizadas da seguinte forma:

Quadro 5 – Função das ferramentas do *WordSmith Tools* e propósito na pesquisa

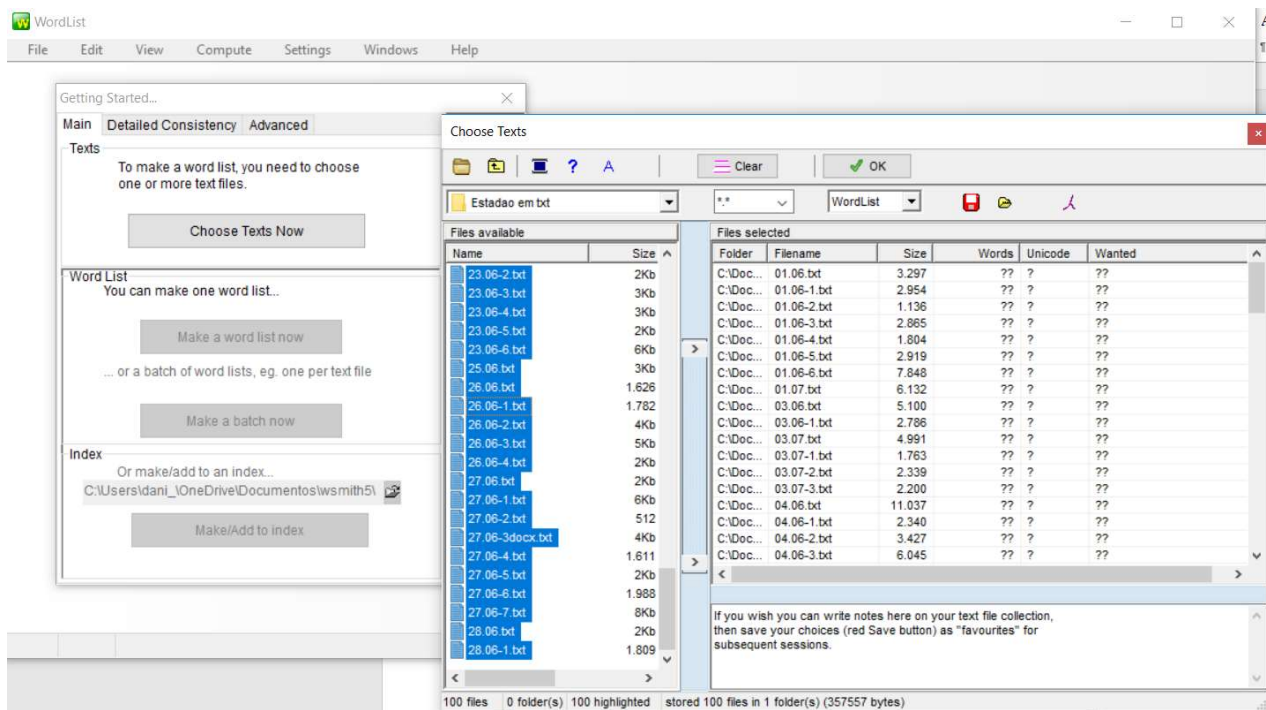
Ferramenta	Função	Propósito na pesquisa
<i>Wordlist</i>	1) Lista por frequência 2) Lista alfabética	Conhecer as palavras mais frequentes
<i>KeyWord</i>	1) Lista de chavicidade de cada jornal 2) Lista de chavicidade dos termos da área de educação em LP e LI	Identificar dentre as palavras mais frequentes no <i>corpus</i> , aquelas que são estatisticamente significativas, na temática estudada, no caso Educação.
<i>Concord</i>	1) Lista de sintagmas	Identificar os sintagmas recorrentes dos termos encontrados.

Fonte: Quadro elaborado pela autora, baseado em Berber Sardinha (2004).

3.1.A lista de palavras

Para utilizar a ferramenta “*WordList*” (lista de palavras), selecionamos todos os artigos de cada jornal escolhido. Após a reunião dos textos, clicando em “*make a word list now*”, (Figura 9) o programa retornou duas listas, uma ordenada por ordem de frequência e outra em ordem alfabética e uma terceira janela com dados estatísticos:

Figura 9 - Seleção dos textos para criação da WordList



Fonte: PrintScreen do programa *WordSmith Tools 6.0*

A Figura 9 apresenta a tela do *WordSmith Tool* no momento da seleção dos textos para a criação das listas de palavras, neste caso os cem artigos do jornal *O Estadão* foram selecionados para a criação das listas de palavras. Observa-se que duas listas foram criadas, sendo uma ordenada pela frequência e outra em ordem alfabética. Nesta pesquisa, utilizamos a lista de palavras ordenadas pela frequência para a comparação com a lista de frequência do corpus de referência (vide seção 2.1.2) e posterior criação da lista de palavras-chave.

A Figura 10 apresenta a janela da lista de palavras do jornal *O Estadão* organizada em ordem frequencial. A coluna **Word** apresenta todas as palavras contidas no *corpus* de 100 artigos originalmente escritos em língua portuguesa. As demais colunas mostram, respectivamente, a frequência, porcentagem total do item no corpus e a quantidade de textos que apresentam determinado item. Como dito anteriormente, a coluna **Lemmas** aparece sem dados, pois não utilizamos a opção de lematização.

Figura 10 - Lista de palavras do jornal *O Estadão* organizadas por ordem de frequência

N	Word	Freq.	%	Texts	%	Lemmas
1	DE	2850	4,95	100	100,00	
2	A	2076	3,60	99	99,00	
3	E	1748	3,04	100	100,00	
4	O	1575	2,73	100	100,00	
5	#	1486	2,58	100	100,00	
6	QUE	1349	2,34	97	97,00	
7	DO	1051	1,82	100	100,00	
8	PARA	908	1,58	98	98,00	
9	EM	804	1,40	97	97,00	
10	DA	785	1,36	99	99,00	
11	COM	638	1,11	95	95,00	
12	É	632	1,10	90	90,00	
13	OS	548	0,95	98	98,00	
14	NO	504	0,88	96	96,00	
15	NA	471	0,82	92	92,00	
16	UMA	470	0,82	89	89,00	
17	AS	461	0,80	99	99,00	
18	UM	445	0,77	89	89,00	
19	NÃO	403	0,70	80	80,00	
20	EDUCAÇÃO	388	0,67	87	87,00	
21	MAIS	338	0,59	80	80,00	
22	SE	333	0,58	77	77,00	
23	DOS	306	0,53	83	83,00	
24	POR	296	0,51	74	74,00	
25	ENSINO	280	0,49	62	62,00	
26	ALUNOS	272	0,47	75	75,00	
27	COMO	269	0,47	80	80,00	
28	DAS	248	0,43	79	79,00	
29	SÃO	232	0,40	76	76,00	
30	ESCOLA	212	0,37	61	61,00	
31	ANOS	181	0,31	53	53,00	

Fonte: PrintScreen do *WordSmith Tools 6.0*

Os *corpora* foram utilizados de maneira individual, portanto criamos quatro listas de palavras, sendo uma do jornal *O Estadão*, uma do jornal *O Globo*, uma do jornal *The New York Times* e uma do jornal norte americano *Los Angeles Times*.

3.2. Palavras-chave

A partir das listas de palavras mais frequentes, a ferramenta *KeyWord* identificou a lista de palavras-chave, ou seja, os termos com maior chavicidade na área da Educação e o programa *WordSmith Tools 6.0* realizou o cálculo estatístico, resultando na chavicidade dos termos. De modo geral, quando a palavra aparece mais vezes no *corpus* de estudo do que no *corpus* de referência, é considerada uma palavra-chave. Este resultado é dado de forma automática por meio do programa de análise lexical *WordSmith Tools 6.0*.

Figura 8 - Lista de palavras-chave do jornal *O Estadão*

Key word	Freq.	%	Texts	RC. Freq.	RC. %	Keyness
EDUCAÇÃO	388	0,67	87	953	0,04	1331,69
ENSINO	280	0,49	62	840	0,04	872,56
ALUNOS	272	0,47	75	1015	0,04	753,34
MÉDIO	157	0,27	43	276	0,01	618,71
JUNHO	114	0,20	91	128		522,81
ESCOLAS	145	0,25	49	363	0,02	492,83
ESCOLA	212	0,37	61	1278	0,05	426,09
CRIANÇAS	156	0,27	39	620	0,03	416,16
COLÉGIO	97	0,17	41	135		415,48
FORMULÁRIO	54	0,09	54	15		331,66
REFORMA	76	0,13	12	145		290,03
CRECHE	41	0,07	10	9		259,41
VAGAS	68	0,12	16	143		249,40
INFERIOR	54	0,09	54	62		245,93
PNE	31	0,05	7	0		231,45
MEC	43	0,07	21	27		229,01
ITINERÁRIOS	33	0,06	8	3		225,88
ESTUDANTES	80	0,14	36	319	0,01	212,91
ANO	138	0,24	59	1135	0,05	211,72
MUNICÍPIOS	47	0,08	14	62		204,89
PROFESSORES	132	0,23	40	1166	0,05	188,57
INFANTIL	47	0,08	26	86		182,31
HABILIDADES	40	0,07	16	46		182,07
ENEM	38	0,07	15	37		181,55
PARA	908	1,58	98	23262	0,99	166,49
HORAS	58	0,10	19	216		160,65
FORMAÇÃO	99	0,17	32	769	0,03	160,32
CURSOS	60	0,10	18	243	0,01	158,18
SISU	21	0,04	5	0		156,78

Fonte: *PrintScreen* do *WordSmith Tools 6.0*

A Figura 11 apresenta a janela com o resultado da comparação entre nosso *corpus* de estudo e o *corpus* de referência *Lacio-Ref* por meio do resultado da prova estatística na coluna *Keyness*, a modo de exemplificação.

Os resultados foram as duas listas de palavras-chave da área de educação, em LP, a lista do jornal *O Estadão* com 228 palavras e a lista do jornal *O Globo* com 241 palavras, totalizando 469 palavras.

Para o levantamento das palavras-chave em língua inglesa, utilizamos como *corpus* de referência, o BNC (*British National Corpus*) que contém 100 milhões de palavras de variadas fontes. Sua última edição foi lançada em 2007. O *corpus* inclui textos de jornais, periódicos, livros acadêmicos, ficção, cartas, memorandos, artigos, entre outros tipos de textos.

Os resultados foram as duas listas de palavras-chave sendo a lista do jornal *The New York Times*, de palavras-chave com 424 palavras e a lista do jornal *Los Angeles Times* com 415 palavras.

Na sequência, realizamos a seleção dos termos para análise. Para tanto, ocupamos dos substantivos comuns como critério de inclusão. Sendo assim, qualquer outra palavra que não fizesse parte da classe de substantivos, foram excluídas, como (artigos, adjetivos, verbos, substantivos próprios, advérbios).

Tabela 1 - Palavras-chave substantivas

Ordem	Palavra-chave	Chavicidade
1	EDUCAÇÃO	1331,69
2	ENSINO	872,56
3	ALUNOS	753,34
4	MÉDIO	618,71
5	JUNHO	522,81
6	ESCOLAS	492,83
7	ESCOLA	426,09
8	CRIANÇAS	416,16
9	COLÉGIO	415,48
10	FORMULÁRIO	331,66
11	REFORMA	290,03
12	CRECHE	259,41
13	VAGAS	249,40
14	INFERIOR	245,93
15	PNE	231,45
16	MEC	229,01
17	ITINERÁRIOS	225,88
18	ESTUDANTES	212,91
19	ANO	211,72
20	MUNICÍPIOS	204,89

Fonte: Tabela elaborada pela autora.

Conforme Tabela 1, os substantivos foram destacados na cor verde, aquelas que permaneceram em branco, não serão utilizadas em nossa análise. Essa seleção foi realizada nas duas listas de palavras-chave em língua portuguesa dos jornais *O Estadão* e *O Globo* e nas duas listas de palavras-chave em língua inglesa, dos jornais *The New York Times* e *Los Angeles Times*.

A seleção dos substantivos do jornal *O Estadão* resultou em uma lista de 159 palavras; do jornal *O Globo*, obtivemos 146 palavras. Dos dois jornais de língua inglesa, *The New York Times* e *Los Angeles Times*, tivemos 206 e 207 palavras respectivamente, conforme Tabela 2:

Tabela 2 - Quantidade de substantivos selecionados

<i>O Estadão</i>	159 palavras
<i>O Globo</i>	146 palavras
<i>The New York Times</i>	206 palavras
<i>Los Angeles Times</i>	207 palavras

Fonte: Tabela elaborada pela autora.

A partir do resultado apresentado (Tabela 2), realizamos uma análise comparativa para compor uma lista única de cada idioma dos termos que se repetiram. As duas listas de cada idioma foram comparadas e as palavras que se repetiram em ambos os jornais formaram uma lista única de cada idioma. Com relação aos termos que se repetiram,

temos como resultado 111 palavras em língua inglesa e 70 palavras em língua portuguesa (conforme Apêndice B e Apêndice C).

A lista dos 70 substantivos em língua portuguesa foi o ponto de partida para análise lexical e análise de correspondência interlínguas. Essa lista foi delimitada, ao passo que restaram 41 termos para análise, já que sinônimos (alunos/estudantes) e palavras com flexão de singular e plural, foram colocadas em um mesmo quadro para análise. Foram retiradas palavras que fazem parte de composições sintagmáticas, como: Básica, Infantil, Curricular), conforme Quadro 6:

Quadro 6 - Termos selecionados para busca de correspondentes

Termos selecionados para correspondência			
1	Alunos	22	Gestão
2	Ano	23	Gestores
3	Aprendizagem	24	Habilidades
4	Aulas	25	Ideia/Ideias
5	Bolsas	26	Infraestrutura
6	Carreira	27	Inscrição / Matrícula
7	Conteúdos	28	Instituição / Instituições
8	Crianças	29	Jovens
9	Currículo	30	Matemática
10	Cursos	31	Municípios
11	Diretores	32	Pais
12	Disciplinas	33	Palavras
13	Diversidade	34	Professores
14	Educação	35	Profissionais
15	Educadores	36	Programa
16	Ensino	37	Redação
17	Escola/colégio	38	Renda
18	Especialistas	39	Sala
19	Evento	40	Semestre
20	Exame	41	Texto
21	Financiamento	42	Vagas

Fonte: Quadro elaborado pela autora.

Conforme Quadro 6, selecionamos 42 termos para análise e busca pela correspondência em língua inglesa. Os termos sinônimos em língua portuguesa, serão colocados em um mesmo verbete e, portanto, apresentam-se juntos, como em escola/colégio, professores/educadores, inscrição/matrícula, assim como termos no singular e plural (ideia/ideias).

3.3. As linhas de concordância

As linhas de concordância apresentam a palavra selecionada para busca com texto antes e depois. Esta ferramenta possibilita que encontremos termos complexos ou composições sintagmáticas a partir de um termo simples.

Nesta pesquisa, utilizamos as linhas de concordância para encontrar as composições sintagmáticas a partir da busca de um termo simples. Nas buscas por **educação** e **ensino**, por exemplo, diversas combinações são encontradas e incluídas nas propostas de correspondência. Apresentaremos a busca pela palavra **ensino** utilizando a ferramenta *Concord* e verificamos 9 combinações diferentes. A figura 12 ilustra uma parte do resultado da busca pela palavra ensino, na qual destacamos algumas das composições sintagmáticas encontradas:

Figura 9 - Busca pela palavra 'ensino' na ferramenta *Concord*



Fonte: *PrintScreen* da tela do *WordSmith Tools 6.0*.

Conforme Figura 12, observando as linhas de concordância da palavra **ensino**, foi possível verificar as composições sintagmáticas ensino superior, ensino médio, ensino fundamental e ensino a distância. Essas composições serão colocadas dentro do verbete **ensino** com seu respectivo correspondente.

3.4. Processo de busca pelos correspondentes

Após a seleção dos termos mais comuns da área da educação em LP e LI e da identificação dos termos que se repetem e que por essa razão podem ser considerados mais frequentes, iniciamos a análise para estabelecer a correspondência terminológica português-inglês. Para tanto, limitamos nossa análise aos 41 termos da lista de termos que se repetem em LP (vide Apêndice D)

O primeiro passo foi organizar as palavras em quadros (conforme quadro 6) divididos em **Língua de Partida (LP)** e **Língua de chegada (LI)**, de forma simples, apenas para a visualização entre uma língua e outra. Para tanto, partimos da direção português-inglês, visando os alunos de idiomas, tradutores e público em geral que buscam o termo no momento da produção em língua estrangeira e por vezes pode encontrar dificuldade por conta de questões culturais e ausência de materiais.

Dessa forma, todas as palavras selecionadas da língua portuguesa, foram organizadas conforme quadro 7:

Quadro 7 - Substantivo em português e correspondente em inglês

Língua de Partida	Língua de chegada
Educação	Education

Fonte: Quadro elaborado pela autora.

Na sequência seguimos as etapas:

- 1) Busca por correspondente Português-Inglês no *corpus* de estudo para atestar a escolha;
- 2) Consulta a ferramenta *Concord* do *WordSmith Tools* também foi utilizada para a busca de possíveis composições sintagmáticas.

- 3) Consulta em dois dicionários bilíngues escolares *Longman Dicionário Escolar* e *Michaelis Dicionário Escolar* para verificar quais as possibilidades de equivalência proposta;
- 4) Consulta ao dicionário monolíngue *Longman Dictionary of Contemporary English*, buscando confirmar a definição em língua inglesa e verificar a correspondência escolhida, bem como, a consulta de sinônimos.
- 5) Consulta ao *Linguee*, disponível *online* ou no formato de aplicativo, que como esclarece Mikhailov e Cooper (2016 p, 202), o “dicionário” *Linguee*, foi destinado, primeiramente e principalmente como uma ferramenta aos tradutores. A busca por uma palavra no *Linguee* resultava apenas na apresentação da palavra de busca e respectivo (s) equivalente (s) inseridos em várias frases e contextos em ambas as línguas contempladas na busca. Assemelhava-se mais a um *corpus* multilíngue que possibilitava pesquisas em *corpus* paralelo;

Segundo Jiménez (2013, p. 12), o fundador do *Linguee* pensou-o como uma alternativa a tradução automática, um serviço de ajuda a qualquer pessoa interessada em produzir traduções de qualidade¹⁷. (tradução nossa)

Ao escrever uma palavra ou grupo de palavras no campo de buscas, o *Linguee* oferece dois tipos de resultados: 1) equivalentes correspondentes à palavra buscada, cada um deles seguido por um exemplo na língua de busca e sua respectiva tradução; 2) frases contendo a palavra buscada dentro de diferentes contextos e sua respectiva tradução.

- 6) Consulta ao site *Education USA*¹⁸, que explica o sistema educacional americano, responde à dúvidas frequentes e oferece um glossário, que apesar de ser monolíngue, apresentando o termo em língua inglesa e posterior definição, nos fornece subsídios para encontrar o melhor correspondente em língua portuguesa.

Abaixo, listamos os procedimentos adotados, elucidando o caminho percorrido para a busca dos correspondentes a partir do termo simples *ensino*:

Ensino

- 1) Correspondência no corpus de língua inglesa;
- 2) Busca do termo por meio da ferramenta *Concord* do WordSmith Tools;

¹⁷ No original: “(...)su fundador, Gereon Frahling, ideó *Linguee* como uma alternativa a la traducción automática, como un servicio de ayuda cualquier persona interesada en acceder a traducciones de calidad.”

¹⁸ Disponível em <<https://educationusa.state.gov/>> Acesso em 07 de março de 2019.

- 3) Busca do termo simples (ensino) nos dicionários bilíngues;
- 4) Busca no dicionário monolíngue;
- 5) Busca de cada composição sintagmática no *Linguee*;
- 6) Definição das correspondências.

A título de exemplificação, o Quadro 8 apresenta os resultados obtidos ao realizar cada uma das etapas descritas acima para chegarmos ao resultado final:

Quadro 8 - Procedimentos adotados durante a análise

Correspondência em língua inglesa no <i>corpus</i> de estudo	<i>education</i>
Busca do termo na ferramenta <i>Concord</i>	<ol style="list-style-type: none"> 1. Ensino privado 2. Ensino Fundamental 3. Ensino Médio 4. Ensino Profissional 5. Ensino Superior 6. Ensino a distância 7. Ensino técnico 8. Ensino Básico 9. Ensino Infantil
Busca no Dicionário Bilíngue <i>Michaelis</i>	<ol style="list-style-type: none"> 1. <i>Teaching, instruction.</i> 2. <i>Education.</i> 3. <i>Training</i>
Busca no Dicionário Bilíngue <i>Longman</i>	<ol style="list-style-type: none"> 1. <i>Teaching. Education. Distance learning; primary education; secondary education; higher education</i>
Busca no Dicionário Monolíngue <i>Longman</i>	<ol style="list-style-type: none"> 1. <i>The process of teaching and learning, usually at school, college, or university.</i> 2. <i>The teaching of a particular subject.</i> 3. <i>The institutions and people involved with teaching.</i> <p><i>COLLOCATIONS</i></p> <ol style="list-style-type: none"> 1. <i>A good education</i> 2. <i>A poor education</i> 3. <i>an all-around education</i> 4. <i>full time education</i> 5. <i>state education</i> 6. <i>public education</i> 7. <i>private education</i> 8. <i>formal education</i> 9. <i>primary/elementar education</i> 10. <i>secondary education</i> 11. <i>college education</i> 12. <i>vocational education</i> 13. <i>pre-school education</i>
	<ol style="list-style-type: none"> 1. Ensino privado: <i>private education</i> 2. Ensino Fundamental: <i>elementary school; primary school; basic education</i>

<p>Busca das composições no <i>Linguee</i></p>	<ol style="list-style-type: none"> 3. Ensino Médio: <i>high school; secondary education</i> 4. Ensino profissional: <i>vocational education; professional education; menos frequentes → professional teaching</i> 5. Ensino Superior: <i>higher education; menos frequentes → higher learning; further education; university education; tertiary education; tertiary level</i> 6. Ensino a distância: <i>distance learning</i> 7. Ensino Técnico: <i>technical education; menos frequentes → technical teaching; technical instruction</i> 8. Ensino Básico: <i>basic education; primary education; menos frequentes: primary school; elementary education; elementary school; primary schooling; basic schooling</i> 9. Ensino Infantil : <i>nada encontrado</i>
<p>Definição das correspondências</p>	<ol style="list-style-type: none"> 1. Ensino privado: <i>Private Education</i> 2. Ensino Fundamental: <i>Elementary School</i> 3. Ensino Médio: <i>High School</i> 4. Ensino Profissional: <i>Vocational Education</i> 5. Ensino Superior: <i>Higher Education</i> 6. Ensino a distância: <i>Distance Education</i> 7. Ensino técnico: <i>Technical Education</i> 8. Ensino Básico: <i>K-12*</i> 9. Ensino Infantil: <i>Early childhood Education</i>

Dessa forma, passamos a discorrer sobre a análise e discussão dos dados, ilustrando os dados resultantes, fornecendo os correspondentes em língua inglesa para os termos da língua portuguesa, bem como informações culturais relevantes ao contexto da pesquisa.

4. ANÁLISE E DISCUSSÃO DOS DADOS

Nesta seção passamos a análise e discussão dos dados deste estudo. A subseção 4.1 discorre sobre a relação de correspondência que propomos a partir da análise dos dados existentes nos dicionários bilíngues, monolíngues e consulta a ferramentas *online*.

A investigação terminológica descritiva, assim como esclarece Aubert (2001) pode desenvolver-se em duas vertentes: a pesquisa pontual e a pesquisa temática. A pontual é aquela que visa a solução de problemas terminológicos isolados de designação, já a pesquisa temática se propõe a levantar o vocabulário terminológico de determinada área ou especialidade.

Dessa forma, nesta pesquisa, propomos a investigação terminológica descritiva temática, sendo a Educação o tema e a apresentar correspondência português inglês para 41 termos, visto a ausência de obras destinadas a esse fim. As informações por vezes, são apresentadas de forma incompleta em dicionários bilíngues escolares, dicionários da língua geral, em glossários monolíngues, porém nunca encontradas em uma obra especializada e bilíngue.

Sendo assim, nesta análise, seguimos alguns pressupostos acerca dos termos funcionalmente equivalentes (TOGNINI, 1996, 2001) e (TOGNINI; MANCA, 2004), atentando-nos aos aspectos léxico-semânticos e tradutológicos. Não visamos uma extensa discussão acerca de tradução, mas buscamos estabelecer relação de correspondência entre os itens mais recorrentes na área de Educação provenientes dos artigos jornalísticos.

Tendo em vista que possuímos dois *corpora*, em línguas diferentes, com diferentes contextos sócio-culturais, assumimos o fato de que a equivalência não pode ser estabelecida apenas no nível da palavra, mas sim, levando em consideração seu contexto.

De acordo com Tognini-Bonelli e Manca (2004), o processo de tradução dá-se em 3 passos:

- I. Identificar uma função específica juntamente com as realizações formais na língua de partida;
- II. Compará-la com outro conjunto (função + realização) formal, ou outros conjuntos na língua de chegada, à luz da fase anterior;
- III. Tentar codificar a função dada em uma realização formal no idioma de destino.

Dessa forma, de posse dos *corpora*, temos como ponto de partida, as palavras com maior chavicidade em nosso *corpus* de estudo, ou seja, aquelas que foram representativas em relação ao corpus de referência (conforme item 3.2). Partimos dos termos em língua portuguesa e buscamos a correspondência na língua inglesa.

Nesse momento da análise, não mostraremos a frequência de cada item, tendo em vista que este requisito foi utilizado no momento da criação das listas de palavras-chave. A ocorrência da palavra nos dois *corpora* de cada língua (dois *corpora* de língua portuguesa e dois *corpora* de língua inglesa) foi nosso requisito para a seleção das palavras para a análise que se segue.

4.1. Análise e correspondência português-inglês

Conforme os procedimentos detalhados na seção 3.4, passamos a discorrer sobre cada substantivo individualmente:

4.1.1. Educação

Língua de Partida	Língua de chegada
Educação	Education

Analisando *educação* como um termo simples, encontramos no corpus de língua inglesa o correspondente *education*. Levando em consideração que educação é utilizado em muitos contextos, verificamos por meio da ferramenta *Concord* quais as combinações encontradas no corpus. Observamos que educação aparece em pelo menos 18 combinações diferentes:

- | | |
|----------------------------------------|-----------------------------------|
| 1. Educação infantil | 10. Educação brasileira |
| 2. Educação de jovens e adultos | 11. Educação inclusiva |
| 3. Educação continuada | 12. Educação especial |
| 4. Educação básica | 13. Educação integral |
| 5. Educação financeira | 14. Educação bilíngue |
| 6. Educação profissionalizante | 15. Educação superior |
| 7. Educação alimentar | 16. Educação artística |
| 8. Educação física | 17. Educação religiosa |
| 9. Educação técnica | 18. Ministério da Educação |

Os dicionários disponíveis atualmente (tomando como base os dicionários consultados nesta investigação)¹⁹ possuem como palavra-entrada de cada verbete os termos simples, então, caso o consulente necessite de alguma destas combinações acima, terá que buscar na entrada *educação* e consultar a microestrutura do verbete.

O dicionário bilíngue Longman Escolar (LONGMAN, 2002) traz apenas a combinação *educação física (physical education)*, assim como faz o *Michaelis Dicionário Escolar* (MICHAELIS, 2009). Ao realizar a busca no dicionário monolíngue, encontramos, após as definições de *education*, um quadro com ‘*collocations*’, em que *education* é utilizado em diferentes combinações.

Com base nas informações encontradas no dicionário monolíngue, nossos *corpora* de estudo e a utilização do *Linguee*, propomos:

Educação infantil → *early childhood education*

Educação de jovens e adultos → *youth and adult education*

Educação continuada → *continuing education*

Educação básica → *basic education* (quando refere-se à educação básica, geralmente trata-se das etapas obrigatórias do ensino brasileiro, compostas pelo ensino fundamental e ensino médio, que serão tratadas no item ‘ensino’)

Educação financeira → *financial education*

Educação profissionalizante → *vocational education*

Educação alimentar → *food education*

Educação física → *physical education*

Educação técnica → *technical education*

Educação brasileira → *Brazilian education*

Educação inclusiva → *inclusive education*

Educação especial → *special education*

Educação integral → *Full-time education*

Educação bilíngue → *bilingual education*

Educação superior → *higher education*

Educação artística → *art education*

Educação religiosa → *religious education*

Ministério da Educação → *Ministry of Education*

¹⁹ (MICHAELIS, 2009; LONGMAN, 2002)

Neste estudo, visamos apenas o levantamento dos correspondentes, portanto não consideramos colocar exemplos e contexto de uso.

4.1.2. Ensino

Língua de Partida	Língua de chegada
Ensino	Education

Assim como educação, ensino também foi observado como termo simples e complexo. Como termo simples, um correspondente possível é *education*. Ao buscá-lo por meio da ferramenta *Concord*, encontramos nove composições sintagmáticas diferentes, a saber:

1. **Ensino** privado
2. **Ensino** Fundamental
3. **Ensino** Médio
4. **Ensino** Profissional
5. **Ensino** Superior
6. **Ensino** a distância
7. **Ensino** técnico
8. **Ensino** Básico
9. **Ensino** Infantil

Organizando as composições de outra forma, temos, as três etapas do ensino básico (ensino infantil, ensino fundamental e ensino médio), concomitante ao ensino médio ou ao término, há o ensino técnico ou profissionalizante; após a conclusão do ensino médio, o ensino superior; aparece também a modalidade ensino a distância e o tipo de ensino privado, conforme Figura 13:

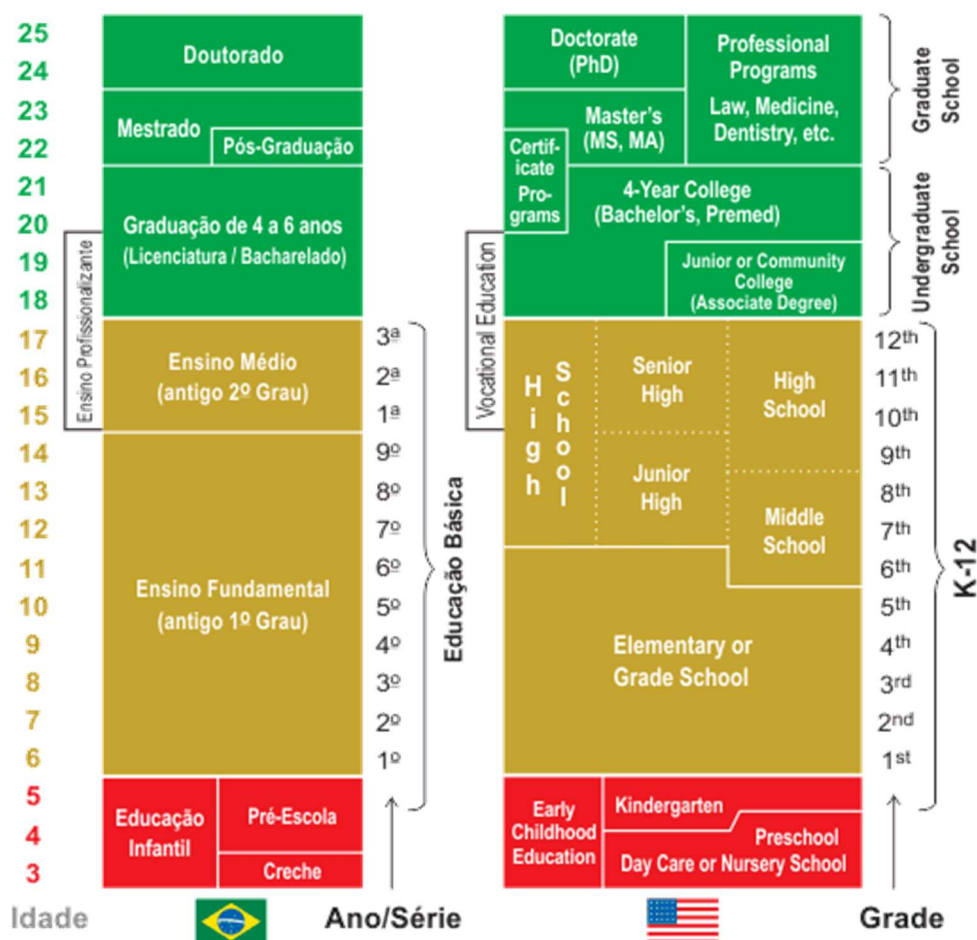
Figura 10 - Reorganização das composições sintagmáticas



Fonte: Figura elaborada pela autora.

Como discutido na primeira seção desta dissertação, os sistemas educacionais brasileiro e americano constituem-se de etapas diferentes. No Brasil, o ensino é obrigatório a partir da pré-escola, nos Estados Unidos, a obrigatoriedade muda de estado para estado. Podemos visualizar a divisão das etapas na Figura 14:

Figura 37 - Comparação dos sistemas educacionais



Fonte: Schütz & Kanomata (2011).

Podemos observar as semelhanças e diferenças dos sistemas educacionais e consequentemente as diferentes designações. No sistema educacional brasileiro, segundo a Lei de Diretrizes e Bases da Educação Nacional, Lei nº 12.796, de 4 de abril de 2013 torna obrigatória a matrícula da criança na educação básica, dividida em 3 etapas (educação/ensino infantil, ensino fundamental e ensino médio) a partir dos quatro anos de idade. Nos Estados Unidos, a obrigatoriedade varia dependendo do estado. Outra

característica refere-se ao ensino médio, no Brasil, é constituído de três anos; nos Estados Unidos, são quatro anos e os estudantes utilizam uma designação para cada um deles (*freshman, sophomore, junior e sênior*), também utilizada durante os anos dos cursos de graduação.

Com base nas informações descritas e na consulta aos dicionário e corpus de estudo, propomos a seguinte correspondência para as composições sintagmáticas:

Ensino privado → *Private Education*

Ensino Fundamental → *Elementary School*

Ensino Médio → *High School*

Ensino Profissional → *Vocational Education*

Ensino Superior → *Higher Education*

Ensino a distância → *Distance Education*

Ensino técnico → *Technical Education*

Ensino Básico → *K-12*

Ensino Infantil → *Early childhood Education*

Conforme nossa proposta de reorganização das composições sintagmáticas (apresentadas na Figura 13), quando nos referimos a educação básica, abarcamos os níveis de educação infantil, ensino fundamental e ensino médio. Em língua inglesa, conforme apresentamos na Figura 14 (SCHÜTZ & KANOMATA, 2011), essa etapa é conhecida como K-12.

4.1.3. Alunos

Língua de Partida	Língua de chegada
Aluno / Alunos Estudante / Estudantes	Student / Students

O dicionário bilíngue *Longman* (LONGMAN, 2002), apresenta *student* como equivalente tanto para aluno, quanto para estudante. Já o dicionário *Michaelis* (MICHAELIS, 2009), apresenta *student* e *pupil* como equivalentes de aluno; e apresenta *student, scholar e learner* como equivalentes de estudante.

Em língua portuguesa, utilizamos aluno/estudante como sinônimos. Nosso *corpus* de estudo de língua inglesa, apresenta *student* como uma das palavras comuns da área de

educação utilizada nos artigos jornalísticos. Sendo assim, buscamos no dicionário monolíngue a palavra *student* e obtivemos definições acerca de cada designação, nas quais, *student* é aquele que estuda em uma escola ou universidade; *pupil* é utilizado na variação britânica, para referir-se a alguém que esteja em uma escola particular ou sendo ensinado por um professor particular; *learner* é destinado àqueles que estão aprendendo uma língua estrangeira.

Dessa forma, consideramos *student* o correspondente adequado para as duas possibilidades da língua portuguesa. Na compilação de um glossário acerca dos termos, seria interessante explicar as outras designações e oferecer contextos de uso.

4.1.4. Escola/Colégio

Língua de Partida	Língua de chegada
Escola / Escolas Colégio / Colégios	School / Schools

Como equivalente às palavras escola e colégio, os dicionários bilíngues apresentam *school*. Em língua portuguesa, tende-se a relacionar a palavra escola ao ensino público, municipal ou estadual, sem que haja necessidade de especificá-la. Já a palavra colégio, tende a relacionar-se à rede privada/particular. Em língua inglesa, essa diferença é dada a partir do uso de um colocado: adjetivo + substantivo. Ex. *public school* (escola pública)

Ao consultar as concordâncias provenientes do corpus de estudo para a palavra escola, encontramos 4 combinações diferentes: escola municipal, escola pública, escola privada, escola estadual, das quais propomos:

Escola municipal → *local school*

Escola pública → *public school*

Escola privada → *private school*

Escola estadual → *state school*

4.1.5. Crianças

Língua de Partida	Língua de chegada
Crianças	Kids

O dicionário *Longman* (LONGMAN, 2002) apresenta *child/children* como única acepção para criança. O *Michaelis* (MICHAELIS, 2009) apresenta *child, infant e baby*. Nenhum dos dicionários bilíngues apresentaram a palavra *kids*, que aparece em nosso

corpus de estudo. No dicionário monolíngue, *kid* é apresentada como a maneira informal para o uso de *child/children*. Neste caso, decidimos por manter a correspondência encontrada em nosso *corpus* de estudo e mais uma vez, na elaboração de um material terminográfico, é interessante esclarecer essas questões de formalidade/informalidade.

4.1.6. Vagas

Língua de Partida	Língua de chegada
Vagas	Admissions

Os dicionários bilíngues apresentam palavras como *vacancy* e *place* como equivalentes à palavra vaga. Em nosso corpus, encontramos *admissions*, ao consultar o dicionário monolíngue, verificamos que *admission* pode referir-se à possibilidade de fazer parte de uma escola. Consideramos o fato de ser aceito/fazer parte em um lugar é o mesmo que ocupar uma vaga.

4.1.7. Ano

Língua de Partida	Língua de chegada
Ano / Anos	Year / years Grade / grades

A palavra ano em língua portuguesa é utilizada em diferentes contextos e assume diferentes sentidos. Refere-se à idade, período/data ou ainda ano escolar. Em língua inglesa, há palavras diferentes para os contextos distintos. Nosso *corpus* apresenta as duas palavras como comuns na área estudada. *Year* refere-se à medida utilizada a designar o período de 365 dias ou doze meses. *Grade* refere-se a um dos doze anos escolares no sistema educacional americano.

4.1.8. Professor

Língua de Partida	Língua de chegada
Professor / Professores	Teacher / Teachers Professor / Professors

Professor é mais uma palavra que sofre divergência semântica na língua inglesa. Em português, a palavra é utilizada tanto nos contextos de ensino fundamental, médio ou superior. Em língua inglesa, no dicionário bilíngue *Longman* (LONGMAN, 2002), já encontramos os equivalentes: *teacher* (para educação básica) e *professor* (para ensino superior). O dicionário *Michaelis* (MICHAELIS, 2009) apresenta os diferentes

equivalentes, porém não distingue o contexto de uso de cada um deles. O dicionário monolíngue (LONGMAN, 1987) distingue os diferentes usos no verbete *teacher*. Portanto, consideramos necessária a distinção das correspondências em uma obra mais detalhada, indicando a que contexto cada um dos correspondentes deve ser utilizado.

4.1.9. Cursos

Língua de Partida	Língua de chegada
Cursos	Degrees

Há em português, a correspondência *course*, apresentada pelos dicionários bilíngues, porém nosso *corpus* de língua inglesa, apresentou *degrees* dentre as palavras-chave, e pensando na possibilidade de uso, optamos por mantê-la como correspondente, já que, podemos utilizar *university degree* como correspondência para curso superior, por exemplo.

4.1.10. Pais

Língua de Partida	Língua de chegada
Pais	Parents

4.1.11. Aulas

Língua de Partida	Língua de chegada
Aula / Aulas	class

4.1.12. Educadores

Língua de Partida	Língua de chegada
Educadores	Educator

Em língua portuguesa, educador vem sendo utilizado como sinônimo de professor. Apareceu em apenas um dos dicionários. O dicionário monolíngue (LONGMAN, 1987) destaca-o como sendo a maneira formal para *teacher*, utilizado principalmente no inglês norte-americano. Outra aplicação da palavra se dá ao contexto de especialistas da área da educação.

4.1.13. Sala

Língua de Partida	Língua de chegada
Sala	Classroom

4.1.14. Matemática

Língua de Partida	Língua de chegada
Matemática	Math

4.1.15. Programa

Língua de Partida	Língua de chegada
Programa	Program

4.1.16. Gestor

Língua de Partida	Língua de chegada
Gestor / Gestores	Administrator / administrators

Segundo o Glossário da Educação²⁰, gestão escolar é a “Organização, mobilização e articulação dos recursos materiais e humanos necessários para assegurar o funcionamento dos estabelecimentos de ensino, assim como a efetiva aprendizagem dos alunos”. Por isso, consideramos que o correspondente *administrator* funciona nesse contexto, visto que aparece em nosso *corpus* de estudo e sua definição no dicionário monolíngue (LONGMAN, 1987) é a de que “alguém cujo trabalho envolve a administração da companhia ou organização”.

4.1.17. Gestão

Língua de Partida	Língua de chegada
Gestão	Administration

4.1.18. Aprendizagem

Língua de Partida	Língua de chegada
Aprendizagem	Learning

4.1.19. Diversidade

Língua de Partida	Língua de chegada
Diversidade	Diversity

²⁰ Disponível em: <<https://www.tuneduc.com.br/glossario-da-educacao/>> Acesso em 10 de dezembro de 2018. A única descrição de autoria que o glossário apresenta é: “A curadoria do conteúdo considerou três referências: “Definições de conceitos para entender inovação em educação” do Porvir, o “Glossário da educação” do Centro de Referências em Educação Integral e o “Anuário Brasileiro da Educação Básica 2017” do Todos Pela Educação e Editora Moderna.

4.1.20. Renda

Língua de Partida	Língua de chegada
Renda	Income

4.1.21. Carreira

Língua de Partida	Língua de chegada
Carreira	Race

Dentre os 42 termos selecionados para análise e busca pela correspondência, somente 21 termos foram encontrados em nosso *corpus* de estudo em língua inglesa. Os outros 21 termos que não se apresentaram no *corpus* foram marcados com a sigla NEC (Não Encontrado no Corpus), conforme Quadro 9:

Quadro 9 - Termos não encontrados no *corpus* de língua inglesa

	Termos não encontradas
1	Bolsas
2	Conteúdos
3	Currículo
4	Diretores
5	Disciplinas
6	Especialistas
7	Evento
8	Exame
9	Financiamento
10	Habilidades
11	Ideia/Ideias
12	Infraestrutura
13	Inscrição / Matrícula
14	Instituição / Instituições
15	Jovens
16	Municípios
17	Palavras
18	Profissionais
19	Redação
20	Semestre
21	Texto

Fonte: Quadro elaborado pela autora.

Conforme observamos no Quadro 9, metade dos termos selecionados para análise e busca por correspondência não foram encontrados em nosso *corpus* de estudo. A razão pode ser devido ao fato de tratarmos de sistemas de ensino distintos, assim como esclarecemos nas seções iniciais desta dissertação. Por exemplo, a divisão escolar brasileira, dá-se por semestres em algumas escolas, então semestre é uma palavra comum

em nosso contexto e talvez não seja no contexto americano, bem como bolsas e financiamento, a citar o FIES (Financiamento Estudantil).

Devido a limitação de tempo e espaço para o desenvolvimento da pesquisa, optamos pela indicação da correspondência nos casos de o termo estar presente em nosso *corpus* de estudo.

Essa seção ocupou-se da análise dos 42 termos selecionados para o estudo segundo os procedimentos descritos anteriormente. Nossa proposta era fornecer um correspondente baseando-nos em nosso *corpus* de estudo, por isso, quando a palavra não foi encontrada no corpus, assinalamos a ausência. Assim como, discorreremos sobre questões divergentes entre a relação de correspondência, quando necessário.

Passamos então a propor a utilização desses resultados para a elaboração de um glossário bilíngue de educação, a ser hospedado em uma plataforma online, conforme detalhamos na seção cinco.

5. PROPOSTA DE GLOSSÁRIO BILÍNGUE

Consideramos que as informações obtidas nesta análise possam fazer parte de um material terminográfico, que vise contribuir com a área e com os usuários que necessitem de tais informações de maneira facilitada e organizada, em um glossário bilíngue, por exemplo.

Um glossário, segundo Welker (2004, p. 25), pretende ser representativo da situação lexical de um único texto manifestado, em situação de discurso exclusivas e bem determinadas. Além de serem extraídas de um corpus delimitado. Vilela (1995, p. 14) define glossário como sendo o vocabulário difícil de um autor, de uma escola ou de uma época. Já Rocha (2010, p.49) retomando Barros (2004, p. 144) define glossário como:

Glossário (termo tolerado: dicionário bilíngue, dicionário multilíngue): pode situar-se tanto no nível do sistema como no da(s) norma(s). Sua principal característica é não apresentar definições, mas tão somente uma lista de unidades lexicais ou terminológicas acompanhadas de seus equivalentes em outras línguas. (BARROS, 2004, p. 144)

Nesta pesquisa, assumimos a definição de Barros (2004) já que objetivamos durante a análise e proposta de correspondência, adicionar apenas o termo e sua correspondência na língua. Porém, há casos em que comentários culturais são cabíveis, portanto, em uma proposta de elaboração de glossário, seriam interessantes de serem incluídos.

O glossário que aborda uma área específica permite objetividade na busca por parte do usuário, visto que dicionários de língua geral, muitas vezes não trazem o termo ou não o apresentam de maneira simples, fazendo com que o consulente tenha que ir de verbete a verbete, na tentativa de sanar sua dúvida. A proposta dessa análise e posterior produção terminológica poderia suprir essas necessidades.

Tendo em vista que a tecnologia está cada vez mais ao alcance de todos, propomos a elaboração do glossário bilíngue através de um site, em que há a possibilidade de criação e hospedagem totalmente grátis. Por meio da plataforma *Webnode*²¹, é possível criarmos um site de maneira simples e fácil. Qualquer pessoa pode criar um site com a *Webnode*. Não é necessário recrutar programadores ou designers ou pagar pela hospedagem do site.

A seguir, passamos à apresentação de como seria o Glossário Bilíngue de Educação, bem como a descrição de sua macro e microestrutura.

²¹ Disponível em: <<https://www.webnode.com.br/>>. Acesso em 20 de fevereiro de 2019.

5.1. Apresentação inicial do Glossário Bilíngue de Educação

A proposta do Glossário Bilíngue de Educação é de um material que seja prático e fácil e que se encaixa perfeitamente em um site. A apresentação inicial do site procura ser clara e concisa, no qual definimos brevemente o que será encontrado neste glossário.

Logo abaixo, há a barra de *links*, que consiste do botão *início*, que volta para a página inicial sempre que acionado; o botão ‘*Sobre mim*’, que apresenta informações sobre a autora; o ‘*contato*’ em que se localiza o *email* para quaisquer dúvidas ou sugestões; e o botão ‘*glossário*’ que leva o consulente ao índice alfabético.

Figura 63 - *PrintScreen* do site Glossário Bilíngue de Educação



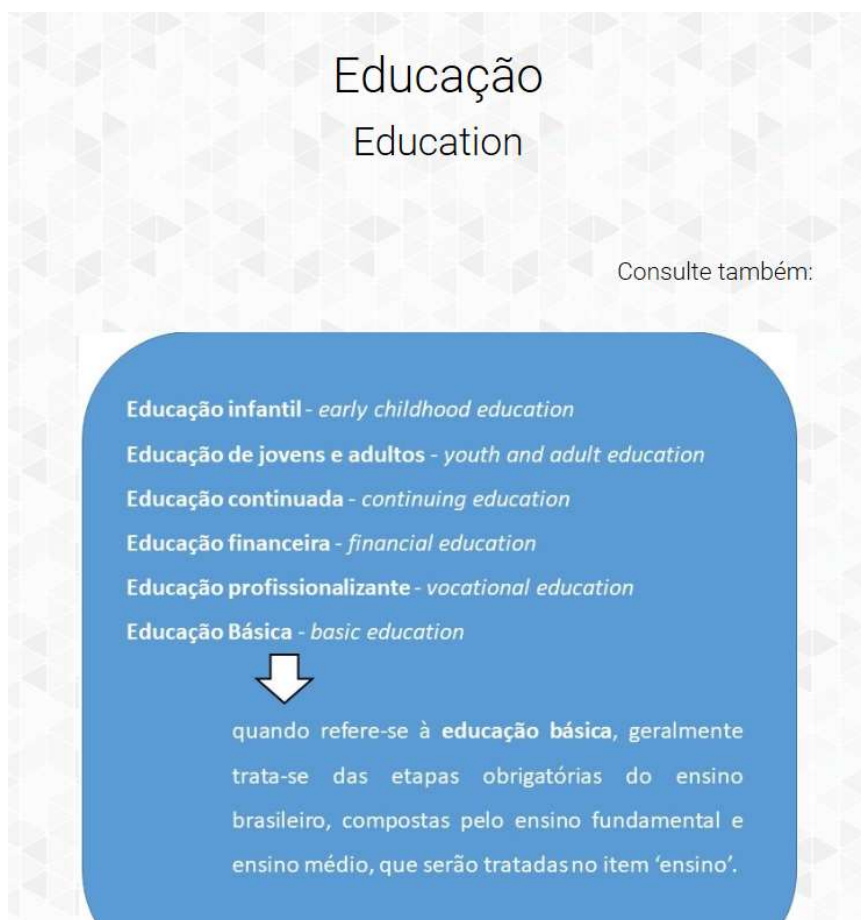
Fonte: Disponível em: <https://vocabularioeducacao.webnode.com/educacao/>. Acesso em: 20 de janeiro de 2019.

Essa apresentação esclarece ao consulente o conteúdo do glossário, bem como a organização da microestrutura, que pode ser encontrada ao clicar em ‘glossário’. Ao invés dos termos já serem apresentados nessa página, há um índice alfabético, possibilitando ao usuário o acesso direto à letra inicial de sua busca, exemplificaremos a apresentação da microestrutura, apresentando o termo ‘Educação’ no Glossário Bilíngue de Educação.

Ao clicar na letra desejada, uma nova página se abre e todos os termos referentes à letra aparecerão um abaixo do outro. Sugerimos uma microestrutura simples, que contenha apenas o termo na língua portuguesa e seu correspondente na língua inglesa. Caso seja interessante, pode-se acrescentar algum comentário explicativo sobre o mesmo.

Ao clicar na letra E, o usuário será direcionado à página dos termos com a letra E e encontrará, Educação, por exemplo. O levantamento dos termos foi feito a partir de termos simples, por isso, Educação será apresentado no topo, com sua respectiva correspondência. Há a frase ‘consulte também’ que indicará ao usuário que há termos complexos, que serão detalhados logo em seguida. Esses termos são apresentados em quadros coloridos, no formato de lista composta pelo termo em português e seu correspondente em inglês, conforme Figura 16:

Figura 83 - PrintScreen da apresentação do termo *educação* no Glossário Bilíngue de Educação



Fonte: Disponível em: <https://vocabularioeducacao.webnode.com/educacao/>. Acesso em: 20 de janeiro de 2019.

Neste caso, há uma informação extra ao usuário sobre o uso do termo ‘Educação Básica’, que geralmente refere-se às etapas obrigatórias do ensino brasileiro, compostas pelo ensino fundamental e ensino médio, que serão apresentadas no verbete do termo ‘ensino’.

Dessa forma, propomos que os dados resultando de nosso levantamento das palavras mais frequentes da área de Educação por meio de artigos jornalísticos possam

ser utilizadas para a elaboração de um material terminológico e em uma plataforma online, atrativa aos usuários.

Passamos então às considerações finais de nossa dissertação, na qual discorreremos sobre o caminho percorrido e lançamos mão de encaminhamentos futuros.

CONSIDERAÇÕES FINAIS

O trabalho que ora se encerra teve como objetivo um estudo teórico-metodológico que possibilitasse o levantamento do léxico mais frequente da área de Educação provenientes de artigos jornalísticos, para que servissem como fonte de dados para elaboração de materiais terminológicos, bem como a proposta de correspondentes na direção português-inglês.

Para alcançar os objetivos propostos, apoiamo-nos no arcabouço teórico-metodológico da LC e no diálogo em que essa área possui com a Terminologia e demais Ciências do Léxico.

Discorremos sobre o tema Educação como fonte para o levantamento do léxico mais frequente, as relações semânticas interlínguas (português-inglês) e as possibilidades de correspondência, caracterizando os sistemas educacionais brasileiro e americano.

Discutimos sobre as possibilidades de aplicações da LC bem como a compilação de *corpora* para promover o alcance dos resultados e o uso de ferramentas computacionais para o auxílio no tratamento de grande quantidade de dados, impossível de realizar-se manualmente.

Discorremos sobre as ciências do léxico, envolvendo as áreas da lexicologia e o estudo da língua geral, bem como as produções lexicográficas, assim como o estudo das áreas de especialidade, por meio do aporte teórico-metodológico da Terminologia. Discutimos sobre as produções terminográficas, com ênfase nos glossários bilíngues, incubida à Terminografia.

Descrevemos em seguida os matérias e métodos que guiaram a coleta dos dados e procedimentos adotados a análise dos dados, descrevendo a ferramenta *WordSmith Tools*, versão 6.0, responsável pela compilação de nossos *corpora* de estudo e posterior criação de listas de palavras-chave, que foram requisitos para o alcance dos objetivos propostos.

Em seguida, passamos a analisar os dados, nos quais, partimos da lista com os 42 termos (substantivos) representativos da área de Educação e a proposta de correspondência dos mesmos. Seguimos a direção português-inglês, visando oferecer uma correspondência semântica aos aprendizes de língua inglesa, tradutores, professores de idiomas e público em geral que procurassem pelos termos da área de Educação no momento de sua produção em língua inglesa.

Sugerimos, na última seção, a proposta de elaboração de um glossário bilíngue de educação com os termos selecionados nesta pesquisa. Propomos um modelo de macro e microestrutura e hospedagem em uma plataforma online, de fácil acesso e que possa contribuir com as produções na área terminológica e contribua para a compreensão e produção de falantes de português na língua inglesa.

Tendo em vista a ausência de materiais bilíngues especializados nessa área, consideramos o levantamento do léxico mais frequente da área de Educação e posterior correspondência na língua, realizada neste estudo, uma fonte viável para a produção de materiais terminográficos, como glossários bilíngues.

REFERÊNCIAS BIBLIOGRÁFICAS

ALTENFELDER, A. H. Reforma do Ensino Médio exige atenção para evitar aumento das desigualdades. **O Estadão**, São Paulo, 04 jun. 2017.

ASTON, G. Applied Corpus Linguistics and the learning experience. In: VIANA, V.; ZYNGIER, S.; BARNBROOK, G. (eds). **Perspectives on corpus linguistics**. John Benjamins, 2013. p. 1-16.

AUBERT, F. H. Tipologia e procedimentos da tradução juramentada. São Paulo: CITRAT/FFLCH/USP, 1998. 2v.

BARBOSA, M. A. *Campo Conceitual e campo lexical dos termos globalização e mundialização: relações*. Revista Brasileira de Linguística, São Paulo, v. 10, n. 1, p. 29-52, 1999.

_____. *Lexicologia, lexicografia, terminologia, terminografia: objeto, métodos, campos de atuação e cooperação*. Estudos Linguísticos, Franca, p. 182 -189, 1991.

BARROS, L. **Conhecimentos de terminologia geral para a prática tradutória**. São José do Rio Preto: NovaGraf, 2007.

_____. *Aspectos epistemológicos e perspectivas científicas da terminologia*. Ciência e Cultura, São Paulo, v. 58, n. 2, p. 22-26, 2006

_____. **Curso Básico de Terminologia**. São Paulo: Editora da Universidade de São Paulo, 2004.

BASTOS, Manoel de Jesus. *Organização do Sistema Educacional Brasileiro*. Revista Científica Multidisciplinar Núcleo do Conhecimento. V. 01, p. 277-286, 2017.

BERBER SARDINHA, T. Pesquisa em Linguística de Corpus com WordSmith Tools. Campinas: Mercado de Letras, 2009.

_____. **Linguística de corpus**. São Paulo: Manole, 2004.

_____. *Linguística de corpus: histórico e problemática*. Delta, São Paulo, v. 16, n. 2, p. 323-367, 2000.

BEVILACQUA, C. R. *Por que e para que a Linguística de Corpus na Terminologia*. In: **Corpora na Terminologia**. TAGNIN, S.; BEVILACQUA, C.. (orgs). São Paulo: Hub Editorial, 2013. p.11-27

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

BIDERMAN, M. T. C. *Terminologia e Lexicografia*. TradTerm: São Paulo. v. 7. 153-181. 2001.

_____. *Dimensões da palavra*. Filologia e Linguística Portuguesa: São Paulo. n.2. p.81-118. 1998.

_____. *O léxico e o vocabulário fundamental*. Alfa: São Paulo. v. 40 27-46.1996.

_____. *A estruturação do léxico e a organização do conhecimento*. Letras de Hoje, Porto Alegre, v.4, n.4, p. 81-96. 1987.

_____. *Teoria linguística: linguística quantitativa e computacional*. Rio de Janeiro: Livros Técnicos e Científicos, 1978.

BOAS, F. *Race, language and culture*. New York: Macmillan, 1940.

BRASIL, Lei de Diretrizes e B. Lei nº 9.394/96, de 20 de dezembro de 1996.

CABRÉ, M. T. *Hacia una teoría comunicativa de la terminología: aspectos metodológicos*. In: _____. **La terminología: representación y comunicación**. Barcelona, IULA, p. 109-127, 1999.

CHOMSKY, N. *Syntactic Structures*. The Hague: Mouton, 1957.

FIRTH, J. R. *Modes of meaning*. In: _____. (Ed.). **Papers in linguistics: 1934-1951**. Oxford: Oxford University Press, 1957.

HALLIDAY, M. A. K. *Corpus studies and probabilistic grammar*. In: AIJMER, K.; ALTENBERG, B. (Ed.). **English corpus linguistics: studies in honour of Jan Svartvik**. London: Longman, 1991.

HARRIS, E. A. *Crime in New York City Schools Is at a Record Low, City Says*. **The New York Times**, New York, 01 ago. 2017.

JIMÉNEZ, E. A. *Linguae y las nuevas formas de traducir*. Skopos, 2013, p. 5-28.

KADER, C. C. C.; RICHTER, M. G. *Linguística de Corpus: Possibilidades e avanços*. **Revista Est. Pesq. Educ. Juiz de Fora**, v. 15, n. 1, jan./jun. 2013.

KRIEGER, M. G. & FINATTO, M. J. B. (2004). **Introdução à Terminologia: Teoria e Prática**. Editora Contexto: São Paulo.

McENERY, T. & WILSON, A. **Corpus Linguistics**. 2nd Edition. Edinburgh: Edinburgh University Press, 2001.

MIKHAILOV, Mikhail; COOPER, Robert. **Corpus Linguistics for Translation and Contrastive Studies: a guide for research**. Abingdon: Routledge, 2016.

NAVARRO, S. L. M. *Glossário bilíngue de colocações da hotelaria: um modelo à luz da Linguística de Corpus*. 2011. 249 f. Dissertação (mestrado) – Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, São Paulo, 2011.

PALMER, H. E. *Second interim report on english collocations*. Tokyo: Kaitakusha, 1933.

PAVEL, S; NOLET, D. *Manual de terminologia*. Tradução de Enilde Faulstich Gatineau, Québec: Direção de Terminologia e Normalização – Departamento de Tradução do Governo Canadense, p. 166, 2001.

RAY, B. D.; Fatos e estatísticas sobre homeschooling nos Estados Unidos e no mundo. In: **National Home Education Research Institute**. 2017. Disponível em: <http://estudosnacionais.com/educacao-domiciliar-estudo-americano/> Acesso em: 10 de janeiro de 2019.

ROCHA, C. F. A tradução juramentada de contratos de compra e venda e de títulos executivos na direção inglês→português: semelhanças e diferenças no uso de termos simples, expressões fixas e semifixas. 2010. 310 f. Tese (doutorado) - Universidade Estadual Paulista, Instituto de Biociências, Letras e Ciências Exatas, 2010.

SANCHEZ, A. Definición e historia de los corpus. In: _____ et al. *CUMBRE: Corpus lingüístico del español contemporáneo: fundamentos, metodología*. Madrid: SGEL, 1995. p. 7 -24

SANTOS, G. M. Proposta de elaboração de glossário terminológico bilíngue para a área de agropecuária – sob a perspectiva da Linguística de Corpus. 2015. 261 f. Tese (doutorado) – Universidade Federal de Santa Catarina, 2015.

SCHÜTZ & KANOMATA. Os sistemas brasileiro, norte-americano e britânico de educação. 2011. Disponível em: <https://www.sk.com.br/sk-edsys.html> Acesso em: 20 de dezembro de 2018.

SCOTT, M. **WordSmith Tools**: version 6. Stroud: Lexical Analysis Software, 2012.

SINCLAIR, J. The search for units of meaning. In: *CICLE de Conferències 95-96: Lèxix, corpus I diccionaris*. Barcelona: Institut Universitari de Linguística Aplicada, Universitat Pompeu Fabra, p. 97107, 1996.

_____. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

STERN, W. **Psychology of Early Childhood up to six years of age**. New York: Holt, 1924.

STUBBS, M. *Text and corpus analysis: computer-assisted studies of language and culture*. Oxford: Blackwell, 1996.

TAGNIN, S. E. O. A Linguística de *Corpus* na e para a tradução. In: VIANA, V.; TAGNIN, S. E. O. (Org.). **Corpora na tradução**. São Paulo: Hub Editorial, 2015. p. 19-56.

TOGNINI-BONELLI, E. **Corpus Linguistics at Work**. Amsterdam/Philadelphia: John Benjamins, 2001.

_____ ; MANCA, E. Welcoming children, pets and guests: towards functional equivalence in the languages of ‘agriturismo’ and ‘farmhouse holidays. TradTerm, São Paulo, 2004, p. 295-312.

_____. Towards translation equivalence from a corpus linguistics perspective. In: SINCLAIR, J. et al. (Ed.). Grammar patterns. London: Collins COBUILD, 1996. p. 197-217.

THORNDIKE, E. L. **Teacher’s wordbook**. New York: Columbia Teachers College, 1921.

VILELA, M. **Ensino da língua portuguesa: léxico, dicionário, gramática**. Coimbra: Almedina, 1995.

WELKER, H. A. **Dicionários – uma pequena introdução à lexicografia**. Brasília: Thesaurus, 2004.

XIMENES, A. P. Homeschooling: como funciona a educação em casa na Flórida. **Brasileirinhos pelo mundo**. 2018. Disponível em: <https://www.brasileirinhospelomundo.com/homeschooling-como-funciona-a-educacao-em-casa-na-florida/> Acesso em: 10 de janeiro de 2019.

ZANETTIN, F. (1998) Bilingual Comparable Corpora and the Training of Translators. Meta: Translators’ Journal, vol. 43, nº 4, p. 616-630.

OBRAS DE REFERÊNCIA

FERREIRA, A. B. H. **Dicionário Aurélio da língua portuguesa**. Curitiba: Positivo, 2010.

MICHAELIS. **Dicionário escolar inglês**. São Paulo: Editora Melhoramentos, 2009.

LONGMAN. **Dicionário escolar inglês-português / português-inglês**. Harlow, 2002.

LONGMAN. **Longman Dictionary of Contemporary English**. Harlow: Longman Dictionaries, 1987.

ANEXO 1 – Artigo do jornal *O Estadão*

Reforma do Ensino Médio exige atenção para evitar aumento das desigualdades

“Em uma sociedade desigual como a nossa, políticas educacionais devem se pautar pela democratização do direito à educação e pela busca de equidade. No caso do Ensino Médio, uma adequada política deve, portanto, buscar garantir que todos os alunos possam ter as mesmas possibilidades de desenvolvimento pessoal, de inserção no mercado de trabalho e de acesso ao Ensino Superior, independentemente de estudarem em tempo de integral ou parcial, ou em turmas do diurno ou noturno.

Garantir a democratização do direito à educação e a equidade é um grande desafio. A nova legislação do ensino médio, porém, contém medidas que podem dificultar avanços nessa direção. Uma delas é a implementação paulatina de escolas de tempo integral. Sabemos – e pesquisas mostram isto – que escolas desse tipo oferecem uma educação de maior qualidade. Elas não podem, no entanto, ser privilégio de alguns. Infelizmente, a legislação corre o risco de favorecer a criação de “ilhas de excelência”, frequentadas pelos estudantes com melhor desempenho acadêmico, resultante sobretudo de seu nível socioeconômico. A pesquisa sociológica e educacional vem desde os 1960 mostrando essa relação entre desempenho e origem social. Pesquisa recente do Cenpec sobre escolas de tempo integral no Ensino Médio evidencia essa relação.

Outra medida que pode dificultar avanços em direção a uma escola mais justa é ligada à flexibilização do currículo. A princípio, trata-se de uma medida bem-vinda, que possibilita uma articulação entre os interesses dos estudantes, seu projeto de vida e o currículo. Entretanto, como o texto aprovado não prevê a obrigatoriedade da oferta de todos os percursos formativos pelas escolas e a possibilidade de escolha poderá não se concretizar para uma parcela importante de nossos estudantes.

Essa limitação deverá ser mais acentuada nos menores municípios. Embora a legislação preveja que percursos formativos possam ser ofertados em parceria com outras instituições, nesses municípios, é menor o número de escolas, bem como são mais restritas as possibilidades de parcerias. É preciso lembrar que Brasil tem quase 1 milhão de alunos do Ensino Médio em municípios de até 20 mil habitantes. ”

ANEXO 2 – Artigo do jornal *The New York Times*

Crime in New York City Schools Is at a Record Low, City Says

“Mayor Bill de Blasio announced on Tuesday that this past school year was the safest for New York City public schools since the Police Department began tracking schools data in 1998, a drop in crime that tracks that of the city overall.

“The 2016-2017 school year was the safest on record in the history of New York City,” Mr. de Blasio said at Middle School 88, the Peter Rouget school in Sunset Park, Brooklyn, where he was flanked by Carmen Fariña, the city schools chancellor, and James P. O’Neill, the police commissioner. “Everyone should be proud of that.”

The Police Department said that arrests fell 8 percent this past school year and summonses dropped by 11 percent. The number of major crimes reported, such as grand larceny and felony assault, also fell from 532 in the 2015-16 school year to 503 in the 2016-17 school year. The city has about 1,800 public schools.

Additionally, Mr. de Blasio said, just two city schools remained on the “persistently dangerous” list, a designation the federal government requires states to make. Just two years ago, there had been 27 city schools in that category. The designation is made using data from a state system called Violent and Disruptive Incident Reporting, or Vadir. Vadir data has not yet been released for the 2016-17 school year.

That did not keep one of the mayor’s consistent foes, the pro-charter group Families for Excellent Schools, from using Vadir data to push a different narrative about violence in the schools. That group said that according to the state data, violence in schools increased in recent years, even as the city data showed it dropping.

The city has frequently criticized the Vadir data as overstating the level of violence in the schools, and others have agreed that it is flawed. The former United States secretary of education John B. King Jr., who was previously New York State’s education²² commissioner, said that the data “rarely reflects the realities of school health and safety.”

Mark Rampersant, the deputy chief executive for safety and security at the city’s Education Department, said Tuesday that the state system counted an episode involving

a first-grader hitting a classmate with an empty milk carton as assault with a weapon. The state recently adopted a new system for tracking school violence, which went into effect in July.

M.S. 88 has embraced a new approach to discipline that encourages schools to move away from suspensions and other punitive measures in favor of what are called restorative justice practices, which include having students discuss their conflicts and misbehavior to try to address problems at their root.

The school even has a dog, named Petey, who comes to school every day to help calm students who are upset and to spend time with those in the guidance office. Petey, an older rescue dog who generally walks the halls in a diaper, is part of the city's Comfort Dog program. He made an appearance at the news conference Tuesday.

In another disciplinary change, the Police Department is testing a new program that gives school safety agents, who are in charge of school security, the discretion to hand out warning cards to students who are 16 or older, instead of issuing criminal summonses for disorderly conduct or possession of small amounts of marijuana. That program is now in 71 schools. Students can still be disciplined by their principals for those infractions.

Although the Police Department said that the number of incidents declined, the number of weapons or "dangerous instruments" found at schools rose last year, from 1,073 in the 2015-16 school year to 1,429 in the most recent school year. But Mr. de Blasio sought to portray that as an indication that the city was getting better at discovering them.

"Look we don't want any weapon in a school, but if a weapon comes in a school, we want school safety to get their hands on it," Mr. de Blasio said, referring to the security agents. "So this begs the question: Over the years, were there a lot of weapons that got into school and never were located and identified? I think that was probably the case."

APÊNDICE A – Título dos artigos jornalísticos selecionados para este estudo

O ESTADÃO
Lições de casa: a construção de uma cultura de estudo
Famílias participam da tarde do macarrão no Glória
The Oscar goes to ...
Projeto: Plugado no mundo adolescente!
Inscrições para o Sisudo segundo semestre se encerram nesta quinta-feira
Telma Vinha explica sobre conflitos e construção de valores durante o desenvolvimento
Justiça quer novo acordo com Prefeito de SP após descumprimento de abertura de vagas em creche
USP propõe 50% de cota para aluno de escola pública até 2021
Em Paraisópolis, falta de opção motivo os imprevistos
Famílias se unem e criam a própria creche
Mindfulness
A intolerância no Brasil e no mundo
Férias “emburrecedoras” ou o ócio criativo
Alimentação saudável: temos fome do quê?
Reforma do Ensino Médio exige atenção para evitar aumento da desigualdade
Ampliar apoio financeiro e técnico às escolas é fundamental, segundo especialistas
Em cidades com 1 só escola, sonho de oferecer 5 ‘futuros’ vira pesadelo
Secretários já cogitam ampliar período do ensino noturno e enviar alunos para cidades vizinhas
Em cidade de SP, docentes cobram reformas físicas
Minas 1,4 mil escolas na fila por reforma física
Reforma do ensino médio esbarra em falta de estrutura e recursos
Ministro Mendonça Filho quer 'liberdade' para Estados definirem a reforma que farão
Colégio realiza atividades pedagógicas diferenciadas na ‘Semana 21’
O papel do brincar na educação
Para sair do papel, política de formação docente precisa ter visão sistêmica e articulação com as políticas de carreira
'A gente olha para a USP e não parece que está no Brasil', diz aluna negra aprovada em 1º lugar
Professora usa xadrez para estimular hábito de leitura em escola
Pela 1º vez, USP aprova cotas raciais e sociais no vestibular
‘Maioria das licenciaturas não vai mudar para o ensino médio’, diz conselheira estadual
Abertas inscrições para bolsas do ProUni do segundo semestre
Depoimento da ex-aluna da Lourenço Castanho e atual aluna da Psicologia da PUC-SP, Bárbara Bim, sobre o IX Congresso ICLOC de Práticas na Sala de Aula.
Plano Nacional de Educação: desafios e perspectivas para a inclusão escolar
Corpo, movimento, cartografia... A geografia na Educação Infantil
Fórum SEE-SAW Panamby
Febre Amarela: Precisamos nos Preocupar?
Xadrez: mais um troféu para a Stance Dual School!
Mordidas e outros conflitos na Educação Infantil: o que fazer?
Plataforma Matific no Ensino da Matemática
Impulsiona: crianças ativas, ágeis e com conhecimento do seu corpo nas escolas brasileiras
Livro que aborda tema do incesto será recolhido de escolas, diz MEC
Papa denuncia tendência a educação elitista: ‘Educar é incluir’
Desigualdade educacional começa na Creche

Inscrições para o ProUni do segundo semestre se encerram nesta sexta
Os tablets e o ambiente escolar: transformação sem volta
Mergulhando na história
Diferentes manifestações culturais na cidade de São Paulo
Aulas imersivas em Inglês por meio de projetos
MEC divulga resultado do ProUni do segundo semestre; confira
Mostra Pequenas Grandes Conquistas
Professor fica menos em escola mais pobre
Para especialistas, problemas em escolas levam a mais desigualdade
Ler, falar, existir: a comunicação, ferramenta essencial da humanidade
Festa Junina agita Colégio Marista Glória
Ensino privado também aceita o ENEM em Portugal
Base Nacional Comum Curricular ainda precisa ser aprimorada
Homens ainda são minoria entre professores de creche e pré-escola
Diversão e aprendizado no Curso de Férias
Lugar de mulher é onde ela quiser, inclusive na Ciência!
Demanda das famílias faz colégios de SP adotarem períodos de até 12 horas
“BatistaComVida”, evento que une solidariedade e diversão recebeu cerca de 1500 Pessoas
Muito mais do que dança – A trama de culturas que enreda o Brasil
Estudantes têm até esta segunda para se inscrever na lista de espera do Sisu
A má educação financeira: uma das mazelas do capitalismo
Alunos da Santi realizam mutirão para revitalizar Praça do Santíssimo Sacramento
Festival Champagnat movimenta toda a Comunidade Escolar
MEC autoriza abertura de 50 novos cursos superiores
‘Estado’ faz cobertura especial e tira dúvidas sobre reforma do ensino médio
4 coisas que você precisa saber sobre o coordenador pedagógico
Docência: profissão ou ofício?
Base Nacional Comum Curricular – Um grande desafio
Quem tem medo do SNE?
O rio Tietê e a escola além dos muros.
Alunos do fundamental já migram mais para EJA do que os do ensino médio
Professor ganha em média R\$ 3.335, diz Inep
Educação: um bem global e de todos
No Rio, professor de História dá aula de Sociologia e o de Artes ensina Inglês
O ensino da língua inglesa no Brasil
Educação abre concurso com 1.878 vagas de diretores de escolas
AGU derruba liminar que representaria prejuízo de R\$ 266 mi para o Fies
Altas habilidades: transformando o desafio em solução
Base Nacional do Ensino Médio é ponto central para mudanças na etapa
Simulação da ONU com alunos do ensino médio
Plano Nacional de Educação completa 3 anos e é preciso avançar no seu cumprimento
Dante Em Foco é imprensa oficial no Greenk
Dante Em Foco realiza cobertura do maior evento de tecnologia e sustentabilidade do País
MEC divulga resultado da lista de espera do ProUni; confira
Nosso convite ao raciocínio
31 quadras poliesportivas aguardam construção em Natal
Educação tem 53% das obras federais paradas ou não iniciadas
Começa nesta terça prazo para inscrição de escolas privadas no Saeb
Reforma aproxima ensino médio brasileiro do exterior, dizem especialistas
Rodas de Encontro com pais e alunos da Educação Infantil
Abaixo a Ditadura!

A pequena galinha ruiva
‘A educação do futuro será a distância’
Depoimento: ‘Em um mês, meu inglês foi do intermediário para o avançado’
Cursos intensivos nas férias ajudam a acelerar aprendizado de idioma; veja opções
Lei autoriza propaganda em uniformes de alunos da rede municipal em São José
Inteligência emocional: por que os alunos precisam ter?
Abertura do Festival Champagnat

O GLOBO
Professores dizem sofrer censura de pais e alunos nas salas de aula
‘O jovem deve conviver com a diferença’, diz educador
Unicef lança plataforma digital para reinserir crianças na escola
Professores podem ganhar até R\$636 trabalhando durante Enem
Pedidos de visto de brasileiros para estudar em Portugal aumentam 148%
Tempos de aprendizado
Pesquisa mostra que 51% dos alunos já usaram celular para estudar
Bolsas do CNPq não serão descontinuadas
CNPq não tem garantia de continuidade do pagamento de bolsas até o fim do ano
Alunos de iniciação científica ficarão sem bolsas do CNPq, diz UFRJ
Formação deve ser voltada para o bem comum, diz Renato Janine Ribeiro
Ameaça de corte no financiamento do CNPq para 100 mil bolsas preocupa cientistas
Um pouco mais de empatia
Novo Fies terá 310 mil vagas, parte delas a juro zero
Estudante com síndrome de Down é aprovado no vestibular da Ufes
Acompanhe ao vivo o evento ‘Educação 360 Tecnologia’
Empatia é indispensável ao abordar direitos humanos no Enem
Educação 360: tecnologia é importante para incluir estudante no processo de aprendizagem
Educação 360: como preparar as escolas para o uso da tecnologia
Educação 360: no painel jovem, adolescentes discutem como usar a tecnologia a favor da educação
Educação 360: ferramentas digitais podem ajudar a conhecer as redes de ensino
Educação 360: ‘Por favor, pense diferente’, pede educador que criou o termo ‘nativo digital’
Educação 360: tecnologia sozinha não promove educação inovadora
Diversidade e excelência
MEC decide recolher das escolas livro infantil que aborda incesto
Editores de livro recolhido pelo MEC afirma que obra passou por ampla avaliação
‘Demonizar um conto é consequência de nossa crise moral’, diz autor de livro recolhido pelo MEC
USP fica em 121º no ranking mundial de universidades
Ensino superior consumirá 58% do orçamento do MEC deste ano
‘Tivemos que fazer isso para não inviabilizar exame’, diz responsável por vestibular da UERJ
Crise leva Uerj a retirar prova de português de exame discursivo
Como usar os simulados para ter um bom resultado no Enem
MEC publica resultados do Prouni para o segundo semestre
Janot inicia ofensiva contra leis municipais que proíbem discussão sobre gênero em escolas
Pesquisa mostra que eleitores preferem líderes autoritários em momentos de crise
Pesquisa mostra que alunos pobres estão em escolas menos preparadas
Seis estudantes vão representar o Brasil na Olimpíada Internacional de Matemática
Rio tem três entre os dez melhores cursos de ensino superior do país
Estudantes que farão Uerj poderão participar de aulas neste sábado
Estudantes contam por que, apesar da crise, vão tentar o vestibular da Uerj

Educação 360 Tecnologia: coragem para mudar e vencer velhas barreiras
Harvard segue como líder em ranking chinês das melhores universidades
‘A escola deve ser o lugar de encontrar soluções para o mundo real’, diz Educador
Entenda como o Enem vai avaliar a sua prova
Centro de estudos facilita a vida de quem vai fazer o Enem
Olimpíada Internacional de matemática começa nesta segunda, no Rio
Países mais corruptos têm desempenho pior no Pisa, diz pesquisa
Escolas buscam meios para preencher lacuna nos livros didáticos causadas por mudanças políticas
Férias é momento de descansar ou manter estudos para o Enem?
Nenhum a menos
MEC pede apoio do Banco Mundial para implementar novo ensino médio
Estudantes que farão Enade podem se cadastrar a partir do dia 14 de agosto
MEC vai aumentar controle sobre avaliação de livros didáticos
Compra de livros didáticos será a cada quatro anos
Escola de Niterói se destaca por ensino inovador do inglês desde a infância
Inscrições para Educação 360 Tecnologia já estão abertas
Prefeituras e estados poderão optar por material didático único
Primeiro ano do ensino médio é a série com maior índice de reprovação e evasão
Unicamp comemora colocação em ranking internacional: ‘orgulho em meio à crise’
Entre 2014 e 2015, 11% dos alunos do ensino médio abandonaram os estudos
‘É necessário avançar para a aprendizagem’, diz representante do Ministério da Educação chileno
Reforma do ensino médio é debatida em evento em São Paulo
Professores do Colorado recebem treinamento com armas de fogo
Coreia do Sul vence Olimpíada de Matemática; Brasil fica em 37º
Delegação da Síria fica de fora da foto da Olimpíada de Matemática
Professora é agredida por aluno após expulsá-lo de sala: ‘Dilacerada’
Enem ainda não tem detectores de metal garantidos para aplicação de prova
Dia de praia para filho de ditador sírio no Rio
Fiocruz: estudantes fazem paralisação contra corte de verba para bolsas
Universidades aumentam o número de cursos de especialização a distância
Estudantes podem solicitar isenção no Enem até domingo
Na reta final para o Enem, aluno contará com material exclusivo
Passatempos que desenvolvem atenção, foco e raciocínio lógico
‘A tecnologia precisa estar na sala de aula’, diz diretor do Google for Education
Projetos de especialização podem beneficiar a sociedade
Escolha precoce de especializações requer alguns cuidados
Ex-alunos do Coppead fazem monitoria em empresas iniciantes
Estudantes buscam pós-graduação de olho na indústria criativa
Universidades oferecem módulos para desenvolver capacidade de liderança
Susan Cain: ‘Introversão não é algo a ser curado’
A conta da crise
Inscrições confirmadas no Enem sobem para 6,7 milhões
Mais de 23% das vagas oferecidas pelo Fies não foram aproveitadas
Erros que derrubam a nota da redação
Alunos da UFMG repudiam projeto de disciplina ‘Casa Grande’ que pedia área para empregados
Especialistas criticam método de ensino que divide alunos por gênero
AGU é contra lei municipal que proíbe discussão sobre gênero em escolas
Educação infantil será incluída na avaliação sobre ensino básico
Falha em inscrição do Sisu afeta até 600 mil estudantes

Pais interferem em escolas que abordam questão de gênero nos livros e vetam conteúdo
Leis municipais que querem barrar discussão de gênero são questionadas no Supremo
Mais de 40% dos alunos têm propensão ao vício em smartphone, diz pesquisa da UFMG
Enem tem 6,1 milhões de inscritos, o menor número desde 2013
‘A comunidade LGBT está sendo expulsa das escolas’, diz ativista pelos direitos humanos
Violência relacionada a identidade de gênero e orientação sexual faz alunos abandonarem escola
Na Tanzânia, garotas são banidas da escola após ficarem grávidas
‘Fundamentalistas construíram poder político na sociedade brasileira’, diz babalaô
‘O jovem precisa ver sentido no que vive na escola’, diz jornalista
Como ficar por dentro das notícias e mandar bem no Enem
Governo autoriza novos cursos de Medicina em 11 municípios

LOS ANGELES TIMES
New report says California is flunking out in producing enough educated adults for the workforce
Officials restore money to L.A. schools soon after making controversial funding cuts
UC Davis chancellor who resigned after ethics probe to return as professor
Brown vetoes a bill to make school districts consider ties between zero-tolerance policies and suicide
Are school suspension rate gaps a problem? ‘Every student is ... unique and Special,’ says DeVos
Schools are boosting graduation rates by offering ‘credit recovery’ But what are Students learning?
Caltech professor resigns after investigation finds he harassed female grad students
Justice Department calls for review of race-based college admissions, alarming civil rights groups
UC Irvine to reinstate all 290 students whose admission offers were withdrawn for transcript problems
Cal State will no longer require placement exams and remedial classes for freshmen
Asian Americans are divided after the Trump administration’s move on affirmative action
Advocacy groups challenging this California state housing rule say it uproots farmworkers’ children from schools
Another record-setting, highlight-filled high school sports season comes to a close
Hate speech vs. free speech: Where is the line on college campuses?
Most private colleges take very few transfers. At USC, about 1,500 get a spot each year
Whittier College president Sharon Herzberger announces her retirement
These small, little-known colleges may be more affordable than you think
Girl with short hair kicked out of soccer tournament: ‘They only did it because I look like a boy’
At 18, this L.A. high school poet often makes audiences go silent
Betsy DeVos is ‘peripherally aware’ that people don’t like her
Approved L.A. Unified settlement would send money to district’s neediest schools
California and 17 other states sue Education Secretary Betsy DeVos for delaying new student-loan protection rules
UC on track to enroll 2,500 more Californians this fall, but admission offers decline from last year’s near-historic gains
New LAUSD board has its first meeting today, with charter school backers in the driver’s seat
A new era begins as pro-charter majority selects Rodriguez to head L.A. school board

San Diego County middle school yearbooks are recalled after a racial slur is noticed on the cover
After 40 years in L.A. schools, this outspoken teacher gives the LAUSD his final rade
Wilmington teen fatally shot near his home dreamed of graduation
Former military school cadet wins \$1.9-million verdict in sexual abuse case
Deportation put on hold for man detained by ICE after dropping daughter off at L.A. School
Cal State L.A. student activist detained by Border Patrol is released
Compton College receives initial accreditation after years of rebuilding
Nick Melvoyn has a lot of ideas for L.A.'s schools
UC Irvine's rare distinction: It's an elite research university that's a haven for Latinos
Mark Zuckerberg's philanthropic group is helping more children see
L.A. school board salaries more than double to \$125,000 a year
Tony winner Ben Platt's L.A. high school days: He shone in different ways than van Hansen, teachers say
Failed a class? A guide to credit recovery in L.A. schools
School's out, now what? 3 ways to keep kids learning during summer break
UC regents panel backs limited use of letters of recommendation at campuses systemwide
Five things to know about L.A.'s school budget
In wake of critical audit, UC regents take a close look at president's budget
Bomb threat at UCLA triggers evacuations to stadium and stirs fear: 'No one really knew what to do'
What exactly is an 'ineffective teacher?' California's definition Doesn't include measures of performance
Former L.A. Unified teacher Rafe Esquith can continue with his lawsuit, judge rules
School board extends Supt. King's contract in current majority's final days
Campus conversation: Maggie White, California State University student trustee
African American students thrive with high graduation rates at UC Riverside
Police report casualties after blast at China kindergarten
Dr. Dre to donate \$10 million for Compton High School's new performing arts center
Does online learning work for LAUSD students taking makeup classes? Study aims to find out
Maryam Mirzakhani, the only woman to win math's Fields Medal, dies at 40 after battle with breast câncer
Trump administration narrows civil rights investigations
Education Secretary Betsy DeVos suspends new rules cracking down on abuses by for-profit colleges
DeVos' new transgender student guidance changes course on bathroom access cases
One child, a \$21-million medical bill: How a tiny number of patients poses a huge challenge for Medi-Cal
An overdose, a Young companion, drug-fueled parties: The secret life of USC med school dean
Drop algebra requirement for non-STEM majors, California Community colleges chief says
In Lynwood, Advanced Placement classes are no longer only for the elite
USC president tries to quell outrage over drug allegations against former medical school dean
Nazarian family donates \$17 million to CSUN's Valley Performing Arts Center
Former USC medical school dean no longer seeing patients; Pasadena police iscipline officer
Cal State trustees grapple with boosting graduation rates and enrollment
Demand for UC immigrant student legal services soars as Trump policies sow uncertainty
Guns kill nearly 1,300 children in the U.S. each year and send thousands more to hospitals
Another jolt for L.A. schools as civic group seeks a role in fixing school district
Alleged conduct by former USC dean 'horrible and despicable' med school head tells angry students

More than 120 layoffs proposed in L.A. school district budget
L.A. Unified approves more spending and layoffs
Two missing Boyle Heights teens found safe
What's Delaine Eastin thinking? The long-shot candidate for governor explains
Police union examines incident at Pasadena hotel involving a former USC dean after an officer is disciplined
Parents of UC Irvine grad who died at Electric Daisy Carnival sue LiveNation and others over his death
Lawsuit alleges hostile environment for Jews on San Francisco State campus
L.A. Unified's new budget, a scholarship from Beyoncé, San Francisco State lawsuit: What's new in education
Lawsuit says North Carolina 'bathroom bill' effects still felt
South El Monte teacher arrested on suspicion of molesting students
L.A. County foster care agency botched many more payments than initially reported
A day before his high school graduation, Redondo High standout guard Ryse Williams dies of cancer at 18
Claremont college suspends students who blocked access to event with pro-police speaker
Senior L.A. Unified official tapped to lead Inglewood schools
Mourning an almost-graduate, picking Betsy DeVos, a sixth-grade food show host: What's new in education
Do charters or traditional schools have it worse? A new study says both
USC received more than a year of questions about former medical school dean's conduct before scandal broke
Texas' bathroom bill debate shows a widening gap between liberal cities and the conservative state
San Diego teacher detained after refusing to answer Border Patrol questions at checkpoint
Supreme Court rules for Missouri church in playground case
LAUSD's newest goal, new charter school research, the mayoral control debate: What's new in education
Brown's budget, Democrats' DeVos concerns, LAUSD spending: What's new in education
Betsy DeVos' actions don't match her words on civil rights enforcement, Senate Democrats claim
Newly elected L.A. Unified board member Nick Melvoin hires a charter school advocate as his chief of staff
1 in 5 L.A. Community college students is homeless, survey finds
Californians, you have until Friday to comment on this key education plan
Students feel safer in more diverse schools, study finds
The politics of math: Is algebra necessary to obtain a college degree?
Melvoin's charter choice, homeless community-college students, raising your voice: What's new in education
New school board member Kelly Gonez isn't as pro-charter as many donors to her campaign
Goodbye, Los Angeles
Former Hacienda Heights basketball coach pleads not guilty to sexually abusing a 14-year-old student
L.A. Trade-Tech álgebra innovations spark uproar — and an internal investigation

THE NEW YORK TIMES
Summer Reading Books: The Ties That Bind Colleges
Star Pupils Offer Advice on Making It in the Arts
Justice Dept. to Take On Affirmative Action in College Admissions
Crime in New York City Schools Is at a Record Low, City Says

Where All the School's a Stage, and the List of Success Stories Is Long
Herma Hill Kay, First Woman to Lead Berkeley Law Faculty, Dies at 82
What Parents Can Do to Nurture Good Writers
6 Things Parents Should Know About Sending Kids Back to School
Why Kids Can't Write
DeVos Abandons Plan to Allow One Company to Service Federal Student Loans
Harassment Suit Against a Stanford Dean Is Rejected
Affirmative Action Battle Has a New Focus: Asian-Americans
How to Conquer the Admissions Essay
'Everyone Deserves an Equal Chance to Get Into College': Readers React
Developmentally Disabled, and Going to College
Going to Hooters and Seeing America
Professors as Targets of Internet Outrage
Klansmen Survive Campus Upheavals
Learning to Learn: You, Too, Can Rewire Your Brain
Listening In on Portland State Activists
Liberal Lessons in Taking Back America
More Diversity Means More Demands
Behind Berkeley's Semester of Hate
Singapore Orders Expulsion of American Academic
Affirmative Action Policies Evolve, Achieving Their Own Diversity
Britain Turns to Chinese Textbooks to Improve Its Math Scores
18 States Sue Betsy DeVos Over Student Loan Protections
De Blasio's Concessions on Charter Schools Are Disclosed
Mark White, Texas Governor Who Championed Public Education, Dies at 77
Cuomo to Give Colleges \$7 Million for Courses in Prisons
An Ice Cream Maker's Latest Recipe: Saving the Local School
Malala Yousafzai, Girls' Education Advocate, Finishes High School
DeVos's Hard Line on New Education Law Surprises States
Prep School Reports Reveal a Chain of Sexual Misconduct
With Snowflakes and Unicorns, Marina Ratner and Maryam Mirzakhani Explored a Universe in Motion
Success Academy Wins Round in Fight Over Preschool Oversight
U.S. to Help Remove Debt Burden for Students Defrauded by For-Profit Chain
Ole Miss Edges Out of Its Confederate Shadow, Gingerly
Ending the Curse of Remedial Math
U.S. Is Investigating Jane Sanders Over Burlington College Bank Loan
Seeing Hope for Flagging Economy, West Virginia Revamps Vocational Track
More Law Schools Begin Accepting GRE Test Results
Reviving a Lost Language of Canada Through Film
Some New York City Students Not Getting Therapy Services, Report Says
A New Kind of Classroom: No Grades, No Failing, No Hurry
A Night of Terror, a Year of Racism
Penn State Student's Dying Hours Play Out in Courtroom Video
'How Far Can They Go?' Police Search of Hundreds of Students Stokes Lawsuit and Constitutional Questions
Rolling Stone to Pay \$1.65 Million to Fraternity Over Discredited Rape Story
What U.Va. Students Saw in Charlottesville
U.S. Halts New Rules Aimed at Abuses by For-Profit Colleges
Harvard President Says She Will Step Down Next Year
New Hampshire Will Investigate St. Paul's School Over Sex Abuse
Connecticut Bill Would Force Fee Disclosures for Teacher Retirement Plans
Education Dept. Says It Will Scale Back Civil Rights Investigations

The Media Brought the Alt-Right to My Campus
A Campus Argument Goes Viral. Now the College Is Under Siege.
New York Schools for Off-Track Students May Face Stricter Rules
Father-Daughter Dance Gets a Makeover in ‘Modern Family’ Era
Harvard’s Graduate Theater Program Halts Admissions for 3 Years
As Paperwork Goes Missing, Private Student Loan Debts May Be Wiped Away
Dropping Out of College Into Life
Wm. Theodore de Bary, Renowned Columbia Sinologist, Dies at 97
A College Built for Canadian Settlers Envisions an Indigenous Future
Dallas Schools, Long Segregated, Charge Forward on Diversity
An Intergenerational Graduación
Consumer Agency Condemns Abuses in Loan Forgiveness Program
Pittsburgh Gets a Tech Makeover
U.S.C. Moves to Fire Former Medical School Dean Over Drug Allegations
Preaching the Value of Social Studies, in a Second Career
Turkey Drops Evolution From Curriculum, Angering Secularists
Does It Matter Who Runs New York City’s Schools?
On Campus, Failure Is on the Syllabus
Why I Bring My Gun to School
He Called the School Board Racist. Now, He’s Joined It.
A Battle Over Prayer in Schools Tests Canada’s Multiculturalism
Marina Ratner, Émigré Mathematician Who Found Midlife Acclaim, Dies at 78
Anti-Chinese Posters at Melbourne Universities Are Tied to White Supremacists
College Students Want to Talk About Sex. They Just Don’t Know How.
Proposal Would Let Charter Schools Certify Their Own Teachers
New York’s Top Court Narrows Suit Seeking More Money for Schools
How Silicon Valley Pushed Coding Into American Classrooms
How Silicon Valley Pushed Coding Into American Classrooms
Cuomo Calls Special Session to Address de Blasio’s Expiring Control of City Schools
School Board Decisions Spur Onondaga Nation Protest
Assembly Approves 2-Year Deal on Mayoral Control of New York City Schools
A New Kind of Tech Job Emphasizes Skills, Not a College Degree
Report Says Diversity Plan Is No Challenge for City
At Troubled City College, President’s Job Remains Unfilled
U.C. Irvine Rescinds Acceptances for Hundreds of Applicants
State Dept. Restores Job Offers to Students After Diplomat Outcry
A College Town Gets Ready for Its Moment Under No Sun
DeVos Is Discarding College Policies That New Evidence Shows Are Effective
Tell Us 5 Things About Your Book: ‘Reading With Patrick’
Your Go-To Guide for Dorm Room Essentials
Sending Sons Off to College, and Finding Solace in a Big Box Store
Campus Political Fights Come Home for the Summer
At Competitive Stuyvesant, Alumni Are Waging a Bitter Contest
Second Phillips Andover Sex Abuse Report Includes Teacher Named by Choate
How to Prepare Preschoolers for an Automated Economy

APÊNDICE B – Substantivos em língua inglesa selecionados para a consulta de correspondência

1	Students	57	Lawmakers
2	School	58	Counseling
3	District	59	Interview
4	Campus	60	Rights
5	Schools	61	Parents
6	College	62	Classroom
7	Charter	63	Law
8	Math	64	Theater
9	Board	65	Performing
10	Student	66	Class
11	Campuses	67	Color
12	State	68	Advocacy
13	Graduation	69	Chancellor
14	Center	70	Members
15	Enrollment	71	Elementary
16	Programs	72	Professors
17	Lawsuit	73	Professor
18	Officials	74	Undergraduates
19	Education	75	States
20	University	76	Justice
21	Program	77	Degree
22	Year	78	Office
23	Federal	79	Leadership
24	Faculty	80	Tenure
25	President	81	Plan
26	Community	82	Educators
27	Website	83	Behavior
28	Admissions	84	Rules
29	Freshmen	85	Applicants
30	Colleges	86	City
31	Nonprofit	87	Families
32	Online	88	Efforts
33	Administrators	89	Diversity
34	Department	90	Learning
35	Gov	91	Attorney
36	Teacher	92	Traditional
37	High	93	Academy
38	Grades	94	Peers
39	Grade	95	Investigations
40	Kids	96	Labor
41	Public	97	Downtown
42	Teachers	98	Hiring
43	Classes	99	Degrees

44	Discrimination	100	Years
45	Academic	101	Achievement
46	Nation	102	Race
47	Universities	103	Children
48	Civil	104	Income
49	Investigation	105	Foundation
50	Statement	106	Decades
51	Administration	107	Goal
52	Freshman	108	Arts
53	Admission	109	Tech
54	Senate	110	Mayor
55	Spokeswoman	111	Terms
56	Summer		

**APÊNDICE C – Substantivos em língua portuguesa
selecionados para compor a lista de termos e respectiva
correspondência**

1	Educação	36	Disciplinas
2	Ensino	37	Superior
3	Alunos	38	Educadores
4	Escolas	39	Inscrições
5	Escola	40	Sala
6	Crianças	41	Matemática
7	Colégio	42	Pública
8	Vagas	43	Colégios
9	Estudantes	44	Privadas
10	Ano	45	Programa
11	Municípios	46	Gestores
12	Professores	47	Curso
13	Infantil	48	Gestão
14	Habilidades	49	Especialistas
15	Cursos	50	Aprendizagem
16	Pais	51	Ministério
17	Aulas	52	Matrícula
18	Anos	53	Evento
19	Nacional	54	Instituição
20	Ideia	55	Inscrição
21	Instituições	56	Semestre
22	Básica	57	Diretores
23	Escolar	58	Ideias
24	Educacionais	59	Redação
25	Estudante	60	Financiamento
26	Aula	61	Estudantil
27	Curricular	62	Conteúdos
28	Jovens	63	Exame
29	Infraestrutura	64	Aprovada
30	Bolsas	65	Diversidade
31	Fundamental	66	Digitais
32	Profissionais	67	Renda
33	Currículo	68	Carreira
34	Educacional	69	Texto
35	Aluno	70	Palavras

APÊNDICE D – Substantivos selecionados para análise lexical e busca pelos correspondentes

Termos selecionados para correspondência			
1	Alunos	22	Gestão
2	Ano	23	Gestores
3	Aprendizagem	24	Habilidades
4	Aulas	25	Ideia/Ideias
5	Bolsas	26	Infraestrutura
6	Carreira	27	Inscrição / Matrícula
7	Conteúdos	28	Instituição / Instituições
8	Crianças	29	Jovens
9	Currículo	30	Matemática
10	Cursos	31	Municípios
11	Diretores	32	Pais
12	Disciplinas	33	Palavras
13	Diversidade	34	Professores
14	Educação	35	Profissionais
15	Educadores	36	Programa
16	Ensino	37	Redação
17	Escola/colégio	38	Renda
18	Especialistas	39	Sala
19	Evento	40	Semestre
20	Exame	41	Texto
21	Financiamento	42	Vagas

APÊNDICE E – Substantivos encontrados no *Corpus* e com proposta de correspondência

Termos selecionados para correspondência	
1	Educação
2	Ensino
3	Alunos
4	Escola/colégio
5	Crianças
6	Vagas
7	Ano
8	Professor
9	Cursos
10	Pais
11	Aulas
12	Educadores
13	Sala
14	Matemática
15	Programa
16	Gestor
17	Gestão
18	Aprendizagem
19	Diversidade
20	Renda
21	Carreira

APÊNDICE F – Substantivos não encontrados no *Corpus* de estudo

	Termos não encontradas
1	Bolsas
2	Conteúdos
3	Currículo
4	Diretores
5	Disciplinas
6	Especialistas
7	Evento
8	Exame
9	Financiamento
10	Habilidades
11	Ideia/Ideias
12	Infraestrutura
13	Inscrição / Matrícula
14	Instituição / Instituições
15	Jovens
16	Municípios
17	Palavras
18	Profissionais
19	Redação
20	Semestre
21	Texto