

UNIVERSIDADE ESTADUAL DE GOIÁS
CÂMPUS CORA CORALINA
PROGRAMA DE PÓS-GRADUAÇÃO EM LÍNGUA, LITERATURA E
INTERCULTURALIDADE

RODRIGO BORGES GONÇALVES

**IDENTIFICAÇÃO DE PADRÕES DE USO DE *PHRASAL VERBS* MAIS COMUNS:
CONFLUÊNCIAS ENTRE LÉXICO, CÓRPUS E INTERCULTURALIDADE**

GOIÁS
2021

RODRIGO BORGES GONÇALVES

**IDENTIFICAÇÃO DE PADRÕES DE USO DE *PHRASAL VERBS* MAIS COMUNS:
CONFLUÊNCIAS ENTRE LÉXICO, CÓRPUS E INTERCULTURALIDADE**

Versão final da dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Língua, Literatura e Interculturalidade da Universidade Estadual de Goiás/Câmpus Cora Coralina, como requisito parcial para a obtenção do título de Mestre em Língua, Literatura e Interculturalidade.

Linha de Pesquisa: Estudos de Língua e Interculturalidade.

Orientador: Prof. Dr. Eduardo Batista da Silva.

GOIÁS
2021

**TERMO DE AUTORIZAÇÃO PARA PUBLICAÇÃO DE TESES E DISSERTAÇÕES
NA BIBLIOTECA DIGITAL (BDTD)**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Estadual de Goiás a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UEG), regulamentada pela Resolução, CsA nº 1.087/2019 sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9.610/1998, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data¹. Estando ciente que o conteúdo disponibilizado é de inteira responsabilidade do(a) autor(a).

Dados do autor (a)

Nome completo Rodrigo Borges Gonçalves

E-mail rborgesgoncalves@yahoo.com.br

Dados do trabalho

Título IDENTIFICAÇÃO DE PADRÕES DE USO DE PHRASAL VERBS MAIS COMUNS:
CONFLUÊNCIAS ENTRE LÉXICO, CORPUS E INTERCULTURALIDADE

Tipo:

Tese

Dissertação

Curso/Programa PÓS-GRADUAÇÃO EM LÍNGUA, LITERATURA E INTERCULTURALIDADE.

Concorda com a liberação documento

SIM

NÃO

¹ Período de embargo é de até **um ano** a partir da data de defesa.

Goiás - GO, 04 de outubro de 2021



Assinatura autor(a)



Assinatura do orientador(a)

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA FONTE

Biblioteca Frei Simão Dorvi – UEG Câmpus Cora Coralina

G635i Gonçalves, Rodrigo Borges.
Identificação de padrões de uso “phrasal verbs”
mais comuns : confluência entre léxico, córpus e
interculturalidade [manuscrito] / Rodrigo Borges
Gonçalves. – Goiás, GO, 2021.
145f. ; il.

Orientador: Prof. Dr. Eduardo Batista da Silva.
Dissertação (Mestrado em Língua, Literatura e
Interculturalidade) – Câmpus Cora Coralina,
Universidade Estadual de Goiás, 2021.

1. Linguística de córpus. 1.1. Itens lexicais -
phrasal verbs. 1.2. Registros - variação linguística.
2. Interculturalidade. I. Título. II. Universidade
Estadual de Goiás, Câmpus Cora Coralina.

CDU: 81'367.625=111(817.3)

Bibliotecária responsável: Marília Linhares Dias – CRB 1/2971

DISSERTAÇÃO DO MESTRADO EM LÍNGUA, LITERATURA E
INTERCULTURALIDADE

FOLHA DE APROVAÇÃO

RODRIGO BORGES GONÇALVES

IDENTIFICAÇÃO DE PADRÕES DE USO DE *PHRASAL VERBS* MAIS COMUNS:
CONFLUÊNCIAS ENTRE LÉXICO, CÓRPUS E INTERCULTURALIDADE

Trabalho aprovado em 04 de agosto de 2021 pela Banca Examinadora constituída pelos seguintes professores:

BANCA EXAMINADORA

Prof. D.r Eduardo Batista da Silva
Universidade Estadual de Goiás
(Presidente)

Prof^a. Dr.a Marília Silva Vieira
Universidade Estadual de Goiás
(Membro interno)

Prof^a. Dr.a Talita Serpa
Universidade Estadual Paulista
(Membro externo)

Dedico este trabalho à minha mãe, Maria Terezinha; ao meu pai, Manoel e a meu irmão Renato. Por sempre acreditarem em meu potencial e pelo apoio e suas palavras de encorajamento nos momentos mais difíceis.

AGRADECIMENTOS

A Deus, pelo dom da vida, por fortalecer a minha fé e me direcionar pelos caminhos da educação e do amor.

Ao meu orientador, Dr. Eduardo Batista da Silva, pela paciência, pelos direcionamentos e, acima de tudo, por nunca deixar de acreditar em mim, mesmo quando a jornada se mostrava difícil.

À Cooperativa de Ensino de Quirinópolis, por ter sempre me apoiado, permitindo que realizasse inúmeras viagens e participasse de inúmeros eventos em busca de conhecimento, mesmo em dias letivos.

A todo corpo docente do POSLLI, pelas inúmeras contribuições linguísticas e, principalmente, pela amizade e dedicação de todos.

À coordenadora do POSLLI e professora de Sociolinguística, Marília Vieira, por acreditar em mim e valorizar o meu aprendizado através da publicação de um artigo.

Ao professor Claude Detienne, pela imensurável contribuição durante a realização dos estágios docentes.

Ao meu pai, Manoel Antunes Gonçalves, por sempre acreditar em mim, pelo apoio financeiro e pelo exemplo de trabalho e de honestidade, os quais foram imprescindíveis para a concretização deste objetivo.

E, principalmente, a minha Mãe, que está no céu, pelo amor incondicional, pela educação transmitida e por ser um exemplo de profissionalismo, meu farol, minha luz, minha eterna professora.

RESUMO

Tomamos como objeto de estudo os *phrasal verbs* (PV), itens lexicais formados pela sequência VERBO + PREPOSIÇÃO ou VERBO + ADVÉRBIO. Tratamos do padrão de uso dos PV, que, pela observação de distribuição em diferentes registros linguísticos, permite um estudo de níveis de formalidade (formal e informal) e modalidade de língua (fala e escrita). Para tanto, faz-se necessário uma abordagem integrada entre léxico, córpus e questões interculturais. Nossa fundamentação teórico-metodológica recorre à Linguística de Córpus (SINCLAIR, 1997, 2004; BERBER SARDINHA, 2004, 2010; BIBER; CONRAD; REPPEN, 2004; ALUÍSIO; ALMEIDA, 2006; OLIVEIRA, 2009; SARMENTO, 2010; MCENERY; HARDIE, 2011; LIU, 2011) e aos preceitos da Interculturalidade (SARMENTO, 2004, CORBETT, 2010; NIETO, 2010; BYRAM, 2012; KRAMSCH, 2013; MELO, 2016). A partir de informações oriundas do *Corpus of Contemporary American English*, o objetivo geral do presente trabalho é descrever o uso dos 150 PV mais frequentes na variante norte-americana da língua inglesa. Os objetivos específicos são os seguintes: 1) identificar a distribuição dos PV em diferentes registros linguísticos; 2) apontar tendência de uso dos PV nas modalidades oral e escrita; 3) apontar tendência de uso dos PV em contextos formais e informais e 4) comparar a ocorrência dos PV em uma perspectiva diacrônica: 2011 (LIU, 2011) e em 2020. No que se refere aos materiais utilizados para o desenvolvimento de nosso estudo, destacamos o *Corpus of Contemporary American English*, Diretrizes Curriculares para o curso de Letras (BRASIL, 2001) e as Diretrizes Curriculares Nacionais para a Formação Inicial de Professores para a Educação Básica (BRASIL, 2020) e, por fim, os dicionários *LONGMAN Dictionary of Contemporary English* e o *Collins COBUILD Advanced Learner's English Dictionary*. Os procedimentos metodológicos compreendem nove etapas, que envolveram leitura dos documentos, consulta à lista de PV (LIU, 2011), buscas e observações no COCA e comparações de uso dos PV. Os resultados indicam que a maioria dos PV analisados encontra-se nos seguintes registros: *fiction* (N=53), *TV/movies* (N=40) e *spoken* (N=21). *Come on* apresentou uma superocorrência na seção *TV/movies* que o elevou ao primeiro lugar considerando as frequências por milhão. Esse fato nos mostra que a linguagem do cinema apesar de representar usos do dia a dia, tem suas frequências artificializadas. *Point out* e *carry out* destacam-se em textos acadêmicos, mas de acordo com os dados, eles despontam em seções diferentes dos textos, pois *point out* é mais recorrente para realçar pontos de vistas de pesquisadores na fundamentação teórica, enquanto *carry out* é mais comum para descrever procedimentos metodológicos. A LC pode contribuir substancialmente para que os dicionários possam esclarecer nuances dos usos do PV em diversos contextos comunicativos.

Palavras-chave: *Phrasal Verbs*. COCA. Registros. Linguística de Córpus. Interculturalidade.

ABSTRACT

We take phrasal verbs (PV) as lexical items, formed by the sequence VERB + PREPOSITION or VERB + ADVERB. We deal with the pattern of use of PV, which, by observing the distribution in different linguistic registers, allows a study of levels of formality (formal and informal) and language modality (speaking and writing). Therefore, an integrated approach between lexicon, corpus and intercultural issues is necessary. Our theoretical-methodological foundation uses Corpus Linguistics (SINCLAIR, 1997, 2004; BERBER SARDINHA, 2004, 2010; BIBER; CONRAD; REPPEN, 2004; ALUÍSIO; ALMEIDA, 2006; OLIVEIRA, 2009; SARMENTO, 2010; MCENERY; HARDIE, 2011; LIU, 2011) and to the precepts of Interculturality (SARMENTO, 2004, CORBETT, 2010; NIETO, 2010; BYRAM, 2012; KRAMSCH, 2013; MELO, 2016) Based on information from the Corpus of Contemporary American English, the general goal of the present work is to describe the use of the 150 most frequent PV in the North American variant of the English language. The specific objectives are as follows: 1) to identify the distribution of PV in different linguistic registers; 2) point out a tendency to use PV in oral and written modalities; 3) to point out the tendency to use PV in formal and informal contexts and 4) to compare the occurrence of PV in a diachronic perspective: 2011 (LIU, 2011) and in 2020. Regarding to the materials used for the development of our study, we highlight the Corpus of Contemporary American English, Curricular Guidelines for the Language course (BRASIL, 2001) and the National Curricular Guidelines for the Initial Training of Teachers for Basic Education (BRASIL, 2020) and, finally, the dictionaries LONGMAN Dictionary of Contemporary English and Collins COBUILD Advanced Learner's English Dictionary. The methodological procedures comprise nine steps, which involved document reading, PV list consultation (LIU, 2011), searches and observations on COCA and PV use comparisons. The preliminary results indicate that most of the PV analyzed are found in the following registers: fiction (N=53), TV/movies (N=40) and spoken (N=21). Come on featured a super hit in the TV/movies section that elevated it to number one considering frequencies per million. This fact shows us that the language of cinema, despite representing everyday uses, has its frequencies artificialized. Point out and carry out stand out in academic texts, but according to the data, they appear in different sections of the texts, as point out is more recurrent to highlight researchers' points of view on theoretical foundations, while carry out is more common to describe methodological procedures. LC can contribute substantially so that dictionaries can clarify nuances of the uses of PV in different communicative contexts.

Keywords: Phrasal Verbs. COCA. Registers. Corpus Linguistics. Interculturality.

LISTAS DE ILUSTRAÇÕES

Figura 1 - Página inicial do COCA	68
Figura 2 - Frequência do PV <i>go on</i> nos registros e nos períodos de tempo	69
Figura 3 - Busca de <i>*-friendly</i>	70
Figura 4 - Ocorrência de [get] [vvn*].....	71
Figura 5 - Fluxograma dos procedimentos metodológicos	78
Figura 6 - Presença de PV por registro no COCA, em 2020.....	79
Figura 7 - PV destaques por registro (LIU, 2011).....	90
Figura 8 – Comportamento de <i>pick up</i> nas subseções da ficção.....	96
Figura 9 – Panorama geral das estruturas léxico-gramaticais de <i>pick up</i>	96
Figura 10 – Tendências de usos de <i>look up</i> nas subseções de <i>fiction</i>	97
Figura 11 – <i>Come in</i> nas subseções de <i>Magazines</i>	99
Figura 12 – <i>Turn out</i> na seção <i>Magazine</i>	101
Figura 13 – <i>Grow up</i> na seção <i>News</i>	102
Figura – 14 As tendências de uso de “ <i>point out</i> ” na seção “ <i>Academic</i> ”	103
Figura 15 – Tendências de uso de <i>carry out</i> no contexto acadêmico.....	104
Figura 16 – Lista lematizada das ocorrências de <i>come on</i>	106
Figura 17 – Frequências de “ <i>come on</i> ” nas subseções de <i>TV/Movies</i>	107
Figura 18 – Frequência de “ <i>What the hell is going on?</i> ”	109
Figura 19 – Frequência da forma contraída “ <i>What the hell’s going on?</i> ”	109
Figura 20 – Frequências de “ <i>get out</i> ” nas seções do COCA.....	110
Figura 21 – Ocorrências de “ <i>gotta get out of here</i> ”	111
Figura 22 – Tendência de uso de “ <i>come up</i> ” na seção <i>Spoken</i>	113
Figura 23 – Frequências de “ <i>End up</i> ” nas seções do COCA.....	115
Figura 24 – Emparelhamentos mais frequentes “ <i>end up</i> ” + <i>ing</i> (<i>END UP*ING</i>).....	118
Figura 25 – “ <i>End up</i> ” + <i>to be</i>	120

LISTAS DE TABELAS

Tabela 1 - Quantidade de tokens nos registros do COCA.....	68
Tabela 2 - Os 10 PV mais frequentes na seção <i>academic</i> do COCA, em 2020	86
Tabela 3 - PV com frequência de uso maior no registro internet	88
Tabela 4 - Os dez PV mais frequentes da variedade norte-americana em 2011 e 2020.....	90
Tabela 5 - Uso de <i>grow up</i> - 2011 e 2020	92
Tabela 6 - Uso de <i>point out</i> - 2011 e 2020	93
Tabela 7 - Uso de <i>carry out</i> - 2011 e 2020.....	94
Tabela 8 - Uso de <i>bring about</i> - 2011 e 2020.....	94

LISTAS DE ABREVIATURAS E SIGLAS

ASL	Aquisição de Segunda Língua
BNC	<i>British National Corpus</i>
BNCC	Base Nacional Comum Curricular
BNC-formação	Base Nacional Comum– formação
BOE	<i>The Bank of English</i>
COCA	<i>Corpus of Contemporary American English</i>
CoEL	<i>Corpus of English Lyrics</i>
DDL	<i>Data Driven Learning</i>
EFL	<i>English as Foreign Language</i>
ESL	<i>English as a Second Language</i>
ESP	<i>English for Specific Purposes</i>
FPM	Frequência por milhão
LC	Linguística de Córpus
LDOCE	<i>Longman Dictionary of Contemporary English</i>
LSWE	<i>Corpus Longman of Spoken and Written English</i>
PV	<i>Phrasal verbs</i>
SydTV	<i>Sydney Corpus of Television Dialogue</i>

SUMÁRIO

1 INTRODUÇÃO	17
2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA	22
2.1 Linguística de Córpus	22
2.1.1 <i>Breve histórico</i>	22
2.1.2 <i>Conceituação</i>	25
2.1.2.1 <i>Córpus</i>	25
2.1.2.2 <i>Linguística de Córpus</i>	30
2.1.3 <i>LC e Ensino</i>	35
2.1.4 <i>Variação linguística nos registros</i>	38
2.2 Interculturalidade.....	51
2.2.1 <i>Conceituação</i>	52
2.2.2 <i>Interculturalidade e Ensino</i>	61
3 MATERIAL E MÉTODO	65
3.1 Material	65
3.1.1 <i>Lista dos 150 Phrasal Verbs de alta frequência (Liu, 2011)</i>	65
3.1.2 <i>Corpus of Contemporary American English</i>	66
3.1.3 <i>Documentos educacionais norteadores</i>	72
3.1.4 <i>Os dicionários</i>	74
3.1.4.1 <i>O Longman Dictionary of Contemporary English</i>	74
3.1.4.2 <i>O Collins COBUILD Advanced Learner's Dictionary</i>	75
3.2 Método	75
3.2.1 <i>Leitura dos documentos</i>	75
3.2.2 <i>Consulta à lista “Frequência e distribuição dos phrasal verbs mais frequentes no COCA”</i>	76
3.2.3 <i>Familiarização com o COCA</i>	76
3.2.4 <i>Busca dos PV no COCA (ocorrência por milhão)</i>	76
3.2.5 <i>Observação da distribuição nos registros</i>	77
3.2.6 <i>Tabulação dos dados</i>	77
3.2.7 <i>Comparação de uso de PV em 2010 e 2020</i>	77
3.2.8 <i>Comparação entre os PV mais comuns: USA x Britain</i>	77
3.2.9 <i>Categorização dos PV (fala x escrita)</i>	77
4 RESULTADOS E ANÁLISE DOS DADOS	79
4.1 PV no COCA.....	79
4.2 Comparação de uso de PV em 2011 e 2020 em termos de frequência e distribuição.....	89
4.3 Tendência de uso dos PV em contextos formais e informais	94
4.3.1 <i>Contextos mais formais</i>	95
4.3.2 <i>Contextos menos formais</i>	105
4.4A interculturalidade e os usos dos PV.	119
5 CONSIDERAÇÕES FINAIS.....	122
REFERÊNCIAS.....	125
APÊNDICE A – Lista de PV no COCA, frequência por milhão, em 2020	129
APÊNDICE B – Gráficos de frequência (COCA 2020)	135

1 INTRODUÇÃO

A presente pesquisa trata do ensino de língua inglesa e direciona suas lentes aos *phrasal verbs* (doravante, PV), unidades lexicais complexas, isto é, formada por um verbo lexical associado a partículas (preposicionais ou adverbiais), PV são polissêmicos, idiomáticos e muito abundantes na língua inglesa, eles costumam gerar dificuldades aos aprendizes do inglês – especialmente, àqueles cujas línguas maternas não possuem equivalentes em seu léxico (LONGMAN PHRASAL VERBS DICTIONARY, 2000; GARDNER; DAVIES, 2007; LIU, 2011).

O arcabouço teórico-metodológico da Linguística de Córpus (doravante, LC) fornece o aporte necessário para que os estudiosos possam capturar e compreender fenômenos da linguagem real, seja em uma perspectiva sincrônica ou em uma perspectiva diacrônica. Além disso, os resultados produzidos por tais pesquisas produzem conhecimento relevante para outros segmentos da sociedade, por exemplo, para o campo do ensino da língua materna bem como de línguas estrangeiras.

Na presente pesquisa, ao associar a LC e Interculturalidade, é possível ampliar os horizontes das investigações, contemplando a relação indissociável entre as formas linguísticas, a construção cultural dos sentidos e a sociedade. Nesse sentido, pode-se produzir conhecimento “da língua”, descrevendo aspectos léxico-gramaticais e semânticos; e “sobre a língua”, relacionando tais aspectos à vida do ser humano, isto é, lançando mão de lentes sócio-culturais. E, por fim, compreendendo que, em um mundo marcado pela diversidade e por mudanças constantes, é necessário formular e reformular os saberes, a fim de se manter sempre atualizado e apto para adequar suas práticas linguísticas aos contextos sociais globalizados do mundo contemporâneo em que as fronteiras geográficas são cada vez mais culturalmente permeáveis.

Quando se trata da formação de docentes de língua estrangeira e das metas que o ensino básico de línguas deve atingir no país, é lugar-comum, tanto em documentos que regulamentam a formação em nível superior (BRASIL, 2002), como na regulamentação do ensino básico (BRASIL, 2018), o desenvolvimento simultâneo da competência linguística e da competência intercultural, mas, na prática, o que se percebe é a reprodução de práticas tradicionais de ensino que fortalecem estereótipos, que não abrangem a pluralidade cultural de corrente de fatores geográficos, sociais e históricos, que se refletem na língua; práticas que apenas se limitam à memorização de vocabulário e à compreensão de regras gramaticais. Portanto, a relação língua-cultura deve ser considerada no ensino de línguas, pois, como afirma Kovalek (2013, p. 785) “desvincular a cultura do ensino de línguas é uma forma empobrecedora de se ensinar língua.”

Além da negligência do componente cultural da linguagem na formação de professores de línguas, a ausência da cultura no material didático também contribui para a proliferação de práticas pedagógicas que reduzem a língua a um conjunto de palavras organizadas por regras sintáticas que

se manifestam em tipos gêneros textuais variados, deixando de lado não só a dimensão cultural, como também desprezando seus aspectos sociais.

O problema não se encerra com o papel secundário do componente cultural no material didático, pois há materiais que tratam a cultura de modo estanque, isto é, encaixotada pelos estados-nações em que se manifesta, as chamadas culturas nacionais, descrevendo, assim linguagem e cultura de forma fixa e homogênea. Portanto, tais materiais contribuem para o desprezo de aspectos dinâmicos da linguagem e da cultura. Acreditamos que o desenvolvimento de falantes realmente capazes de lidar com a linguagem em seus mais variados contextos, eficientemente, requer tanto material didático coerentemente desenvolvido para este fim, como também profissionais docentes igualmente capacitados.

Vários estudiosos vêm realizando pesquisas cujos objetos são os PV, seja para a descrição linguística (BIBER *et al.*, 1999; GARDNER; DAVIES, 2007; DAVIES; GARDNER, 2010; LIU, 2011), seja para produção de conhecimento pedagógico (RAMPAZZI, 2008).

Merece destaque um aspecto que, embora relevante, passa despercebido aos olhares de desenvolvedores de materiais de ensino, de professores e de aprendizes: o comportamento dos PV em contextos formais e informais, orais ou escritos. Maior atenção deve ser direcionada para o comportamento linguístico dos PV, especialmente, um olhar detalhado sobre quais lexias são mais usadas em registros/gêneros da oralidade e quais são mais presentes na escrita. Acreditamos que essa informação pode contribuir não apenas para a formação de professores de língua inglesa, como também, pode ser um relevante aporte para o desenvolvimento de material didático baseado não apenas em textos autênticos e vocabulário mais frequente, mas, que acima de tudo englobe aspectos sócio-culturais da linguagem.

Cumpramos destacar que trataremos gêneros textuais como registros, em consonância com o pensamento de Biber e Conrad (2009). Nessa perspectiva, traços léxico-gramaticais são tratados de forma funcional, isto é, seus usos estão ligados aos propósitos comunicativos e são influenciados pelo contexto situacional dos textos. Logo, é importante salientar que, ao lançar mão do termo “registros”, excede-se as fronteiras estabelecidas pela compreensão de registro formal ou registro informal, comumente empregados por outras áreas de estudos da linguagem. Ao tratarmos a linguagem sob uma perspectiva de registro, destacamos traços linguísticos quantificáveis, isto é, cujos aspectos podem ser descritos por meio da frequência, da ocorrência e da co-ocorrência.

O conhecimento acerca do comportamento desses PV pode orientar os responsáveis por desenvolver materiais didáticos, na difícil tarefa de escolher quais devem ou não compor o programa de ensino, já que considerando 150 um número muito grande, o comportamento em determinados registros da linguagem pode apontar quais se destacam no tipo de texto em estudo, ou, que ainda sejam mais recorrentes em determinadas estruturas léxico-gramaticais, tornando a

decisão pautada em critérios quantitativos para fins pedagógicos, ao invés de baseá-la apenas na intuição.

Uma vez que, dominar uma língua vai muito mais além de conhecer seus aspectos semânticos e léxico-gramaticais, é fundamental que o professor formador tenha consciência também dos domínios comportamentais, não apenas relacionados a registros que serão destinados ao ensino da língua, mas, acima de tudo, é preciso dominar aqueles conhecimentos diretamente utilizados no seu cotidiano, como ao divulgar os resultados de pesquisas acadêmicas, uma tarefa fundamental para um professor-formador, no âmbito da academia.

Nossa pesquisa diferencia-se das anteriores por buscar atualizar os dados de pesquisa realizada há mais de dez anos e direciona suas lentes científicas para produzir dados que podem auxiliar tanto na formação de novos professores da área de Letras quanto no fornecimento de dados para a descrição da língua inglesa. Além disso, promove discussão acerca de aspectos linguísticos acerca dos PV, considerando suas tendências de uso em uma perspectiva diacrônica e sócio-cultural.

Diante dos problemas tratados até aqui, com intuito de promover discussão acerca do ensino de línguas, em especial ao de língua inglesa, considerando as expectativas definidas pelos órgãos responsáveis pelo ensino de línguas no Brasil, a relação língua-cultura e as novas demandas de ensino de línguas no contexto da globalização, traçamos as seguintes perguntas de pesquisa:

- 1) Quais são as expectativas dos documentos oficiais (BRASIL, 2002, 2020) relacionadas à formação de professores quanto ao conhecimento dos PV?
- 2) Qual é a relação entre a Interculturalidade e o uso de PV?
- 3) Qual registro configura-se como um lócus de ocorrência de PV?
- 4) Como é o tratamento dos PV em dicionários, considerando os padrões de uso evidenciados pelo COCA?

No Brasil, com relação aos PV, é importante enfatizar o trabalho de Rampazzi (2008), que lançou mão de análises baseadas em *cópus* para identificar PV de alta frequência, bem como as estruturas léxico-gramaticais recorrentes associadas a eles.

No exterior, podemos citar os trabalhos realizados por Biber *et al.* (1999), cujos autores elencaram uma lista de 31 PV de alta frequência, utilizando dados extraídos do *Corpus Longman Spoken and Written English* (LSWE). A pesquisa realizada por Gardner e Davies (2007), que resultou em uma lista dos 100 PV mais frequentes da língua inglesa em sua variedade britânica, cujos dados foram extraídos do *British National Corpus* (BNC), um *cópus* composto por mais de cem milhões de palavras. Também merece destaque a lista dos 100 PV mais frequentes da variedade norte-americana, obtidos por Davies e Gardner (2010), por meio de uma pesquisa baseada no *Corpus of Contemporary American English* (COCA).

Com o objetivo de descrever o uso dos PV em uma perspectiva sócio-cultural e intercultural de ensino da língua inglesa, por meio de uma abordagem baseada em *córpus*, este trabalho almeja identificar usos de PV em 8 registros da base do COCA, associando seus usos à fala e à escrita, em contextos formais e informais. Acreditamos que o conhecimento produzido nesta pesquisa pode contribuir significativamente para a superação dos problemas apontados.

Nesse contexto, nossa pesquisa produz resultados que podem não só preencher lacunas presentes no âmbito da LC, como também pode contribuir para aqueles envolvidos direta ou indiretamente com o ensino e aprendizagem de língua inglesa. Acreditamos que os resultados da pesquisa podem contribuir de modo significativo para a sociedade no âmbito geral, uma vez que, ao direcioná-los para o professor formador, no ensino superior, produzir-se-á um efeito em cadeia alcançando os professores em formação, os quais, ao adentrarem no mercado de trabalho, podem contribuir para a formação de usuários da língua inglesa.

Para a presente pesquisa, tomamos como ponto de partida o trabalho de Liu (2011), que gerou dados complementares para as pesquisas mencionadas, anexando os dados obtidos por Biber *et al.* (1999) ao trabalho de Gardner e Davies (2007), além de realizar novas buscas, resultando em uma nova lista de 150 PV que abrange as duas variedades da língua inglesa, a saber: a britânica e a norte-americana.

A partir das informações oriundas do COCA, o objetivo geral do presente trabalho é descrever o uso dos 150 PV mais frequentes na variante norte-americana da língua inglesa. Os objetivos específicos são os seguintes: 1) identificar a distribuição dos PV em diferentes registros linguísticos; 2) apontar tendência de uso dos PV nas modalidades oral e escrita; 3) apontar tendência de uso dos PV em contextos formais e informais e 4) comparar a ocorrência dos PV nos anos de 2011 (LIU, 2011) e em 2020.

Uma vez que os PV são recorrentes tanto na modalidade falada quanto na modalidade escrita, é relevante, não só indicar quais são os mais frequentes, como também demonstrar quais possuem mais uso em situações de fala e escrita e em contextos informais e formais – produzindo aporte para o desenvolvimento da competência linguística e intercultural dos aprendizes da língua inglesa.

O público-alvo do presente trabalho é constituído por professores-formadores, tornando-os conscientes de aspectos inerentes à língua inglesa: alinhando suas práticas às exigências presentes em documentos norteadores oriundos de instituições responsáveis pela regulamentação do ensino superior, específica ao curso de Letras no país (BRASIL, 2002) e alinhando, por conseguinte, as práticas dos professores da Rede Básica, os quais são formados nos cursos de Letras.

Ao direcionar nosso estudo para o profissional responsável pela formação docente, almejamos vincular prática e teoria, ao passo que disponibilizamos informações úteis tanto para o corpo docente, que absorve esse novo conhecimento, quanto para o pesquisador acadêmico que pode se

beneficiar desse saber para disseminar os resultados de suas pesquisas, lançando mão de uso da língua inglesa de forma precisa e autêntica.

O presente trabalho será estruturado em 05 seções, a saber: além dessa introdução, o capítulo 2 que se destina ao detalhamento de nossa fundamentação teórica. O terceiro capítulo apresenta os materiais e o caminho metodológico percorrido. O quarto capítulo destina-se à discussão dos resultados encontrados, enquanto, na seção final, constarão nossas considerações finais. Por último, incluímos nossas referências bibliográficas, apêndices e anexos.

2 FUNDAMENTAÇÃO TEÓRICO-METODOLÓGICA

A fundamentação teórica recorre à Linguística de Córpus e à Interculturalidade.

2.1 Linguística de Córpus

Nas próximas subseções, apresentamos um breve histórico da área, uma conceituação de córpus e de Linguística de Córpus, relação entre Linguística de Córpus e ensino e, por fim, a questão da variação linguística nos registros.

2.1.1 Breve histórico

O primeiro córpus eletrônico foi criado no ano de 1964, o Córpus *Brown*, uma coletânea de 1 milhão de palavras – o que representou um grande marco.

As pesquisas nesta área se iniciaram em meados do século XX, momento em que a corrente linguística chomskyana surpreendia a comunidade científica com teorias relacionadas à linguagem que satisfaziam os olhares dos pesquisadores da época, mas que careciam de um embasamento que representasse a linguagem natural. Sua teoria foi edificada não sobre amostras linguísticas que capturavam a linguagem em seu curso, mas sim, trabalhava com textos artificiais, isto é, criados (pela mente humana, ou melhor, pela mera ilusão intuitiva) para propósitos científicos da Linguística Gerativista.

A LC vivenciou embates com a corrente da linguística racionalista, uma vez que seu arcabouço teórico-metodológico baseava-se em descobertas científicas alicerçadas em estudo de amostras reais da linguagem, isto é, por meio de métodos empíricos.

Antes da compilação do córpus Brown, outros *corpora* surgiram para satisfazer inúmeras necessidades de pesquisas ligadas não apenas à necessidade de descrever fenômenos linguísticos, mas também para produzir conhecimento acerca da linguagem no que diz respeito ao ensino. Nesse sentido, podemos destacar os trabalhos que trouxeram à luz listas de palavras de alta frequência da língua inglesa, primeiro com Thorndike e mais tarde com West em 1953. Segundo Berber Sardinha (2000, p. 326), esta seja “talvez a mais famosa descrição do léxico inglês pré-computador. A pesquisa de West dá detalhes do que seria as 2 mil palavras mais frequentes do inglês”. O prestígio em torno da realização desses estudos baseados em *corpora* impulsionou a compilação de amostras cada vez maiores, mais abrangentes e diversificadas, utilizadas por diferentes áreas da Linguística, para finalidades variadas, contribuindo com descrições linguísticas para satisfazer necessidades de cunho científico, comercial (dicionários e gramáticas) e pedagógico. Os próximos parágrafos trazem exemplos de algumas dessas realizações.

O *The Bank of English* (BoE) é parte do *Collins Word Web*, uma coleção de aproximadamente 645 milhões de palavras de inglês falado e escrito, desenvolvido para o estudo da língua em uso. Tem uma ampla variedade de tipos de textos, isto é, registros de escrita e fala de centenas de fontes diferentes. Seu banco de dados é atualizável e parte de amostras coletadas em 1990. Aproximadamente 40 % das fontes são britânicas, enquanto 30% são de amostras do inglês americano, e os outros 30% correspondem a amostras da Austrália, da Nova Zelândia e do Canadá. Ele possui amostras de linguagem falada informal representada por gravações de conversas casuais, entrevistas, reuniões e transcrições de programas de rádio e TV. Todo esse rigor em sua compilação, já evidencia que se trata de um *corpus* utilizado para fins lexicográficos, isto é, para a produção de dicionários.

Em 1994, conclui-se um trabalho audacioso que culminou na concretização do BNC (*British National Corpus*), o primeiro a bater a marca de 100 milhões de palavras. Sua coleção de amostras representa a variedade britânica da língua inglesa, é composto por textos de fala (10%) e da escrita (90%). Foi projetado para ser uma amostra do inglês britânico do final do século XX. Um exemplo de trabalho realizado através de suas amostras é o realizado por Gardner e Davies (2007) que apresentou dados referentes à frequência de 100 PV do inglês britânico. A identificação de unidades gramaticais próprias dos PV foi facilitada devido ao fato da amostra do BNC ser etiquetada morfologicamente.

O COCA, desenvolvido em 2008, destacou-se por ser o primeiro *corpus* linguístico projetado para permitir pesquisas tanto sincrônicas como diacrônicas relacionadas ao inglês em sua modalidade norte-americana, pois trata-se de um *corpus* dinâmico (os tipos de *corpus* serão detalhados na próxima seção e a análise detalhada do COCA será apresentada na metodologia desta dissertação) que tem possibilitado a realização de inúmeras pesquisas sobre a linguagem presente em sua coletânea. É importante dizer que além de um *corpus* monitor, ele se destaca por estar disponível no ambiente da *web*, no *site* <https://www.english-corpora.org/coca/>, portanto é fácil acessar o seu banco de dados para a realização de estudos diversos.

Atualmente, vários *corpora* são compilados por pesquisadores que desenvolvem pesquisas específicas relacionadas a diversas línguas, em diversas partes do mundo, como por exemplo, o *Sydney Corpus of Television Dialogue* (SydTV), projetado por Bednarek (2018) para permitir estudos acerca da linguagem da TV, como uma variedade linguística. Segundo a própria autora, “está claro que há um escopo para pesquisas linguísticas que examinam uma ampla variedade de narrativas de TV, permitindo um novo foco sobre os diálogos

contemporâneos de TV, como uma variedade linguística.¹” (BEDNAREK, 2018, p. 82, tradução nossa), ela ainda acrescenta, ao mencionar a relevância do *corpus*, “SydTV foi projetado para preencher essa lacuna, com o objetivo de capacitar diferentes tipos de análises linguísticas.²” (BEDNAREK, 2018, p. 82, tradução nossa)

No Brasil, inúmeras pesquisas também são realizadas, desde o lançamento da obra *Linguística de Corpus*, publicada por Berber Sardinha (2004), responsável pela divulgação da LC no país e um pesquisador com produção respeitável na área. Mencionar ou deixar de mencionar trabalhos realizados no âmbito da LC no Brasil é, certamente, correr o risco de cometer injustiças, uma vez que há produções em inúmeras universidades brasileiras, então, apenas a título de exemplificação e escolha, optamos por destacar o trabalho de Silva e Babini (2011) por compilarem um *corpus* de língua inglesa composto por amostras linguísticas em diversas áreas de especialidade com o objetivo de demonstrar a relevância da LC na preparação de material terminológico ao analisar aspectos linguísticos presentes em terminologias da área das Redes Neurais Artificiais.

Aqui, exemplificamos apenas algumas coletâneas de textos que vêm sendo compiladas desde o surgimento da LC até os dias atuais, para possibilitar as pesquisas baseadas nos princípios dessa área de estudos linguísticos que paulatinamente vem inovando suas metodologias, não só apenas impulsionada pela evolução da tecnologia, mas principalmente por meio da criatividade de seus pesquisadores.

Ao traçar uma pequena síntese da história da LC nos últimos 70 anos, optamos por apresentar algumas pesquisas associadas a vários *corpora* compilados para servirem de amostras da linguagem em diferentes empreitadas científicas. O *corpus* foi fundamental para demonstrar que a linguagem tem natureza probabilística, diferentemente do que defendiam alguns linguístas gerativistas, ao insistirem em dizer que o sistema linguístico podia produzir infinitas possibilidades, porém ao comparar alguns padrões (artificiais) criados apenas para estudos linguísticos com amostras linguísticas reais, extraídas de *corpora*, ficou claro que, apesar de possíveis, elas não ocorriam no uso cotidiano da linguagem natural.

Engana-se quem acredita que a LC, desde seus primórdios, lida com amostras gigantescas da linguagem representadas por coletâneas de textos formadas por milhões de palavras. Conforme apresentado aqui, na verdade, trata-se de uma área que tem evoluído, ao passo que a tecnologia oferece ao mundo computadores cada vez mais potentes e ferramentas

¹It's clear that there is a scope for linguistic research that examines a wider range of TV narratives, allowing a new focus on contemporary television dialogue as a language variety.”

²(SydTV) was designed to fill this gap, with the aim of enabling different types of linguistic analysis.

linguísticas mais sofisticadas, as quais permitem aos linguístas de hoje investigar amostras gigantescas da linguagem em um espaço de tempo muito curto, algo antes impensável.

2.1.2 Conceituação

Na presente seção, trazemos a conceituação de *cópus* e, na sequência, a conceituação de LC.

2.1.2.1 *Cópus*

Iniciamos a apresentação das definições com uma das explicações mais conhecidas na área, que é a seguinte:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise (SANCHEZ, 1995, p. 8-9 apud BERBER SARDINHA, 2012, p. 321-322).

Cópus, também pode ser explicado por meio das palavras de Berber Sardinha (2000, p. 325) como “conjunto de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística.”; ou seguindo a perspectiva de Viana (2010, p. 24), *cópus* é entendido como “uma compilação eletrônica e criteriosa de (amostras de) textos que ocorrem naturalmente com o objetivo de representar uma dada língua ou algum de seus aspectos mais pontuais de forma a possibilitar uma análise linguística previamente delineada”; ou de acordo com Hunston (2002, p. 2 apud SARMENTO, 2010, p. 90-91), um *cópus* é uma “coletânea de exemplos naturais de linguagem, que consistem desde algumas frases até conjuntos de textos escritos ou gravações orais que foram coletados para serem usados como base para pesquisa linguística.”; ou, nas palavras de McCarten (2007, p. 2, tradução nossa), “um *cópus* é basicamente uma coleção de textos que é armazenada em um computador. Os textos podem ser de linguagem falada ou escrita. [...] É então, possível analisar a linguagem em um *corpus* com *softwares* para ver como as pessoas realmente falam ou escrevem.³”; ou, de um modo mais sintetizado, pode ser

³A *corpus* is basically a collection of texts which is stored in a computer. The texts can be written or spoken language. [...]It is then possible to analyze the language in the *corpus* with *corpus* software tools to see how people really speak or write.

definido de acordo com Biber, Conrad e Reppen (2004, p. 12, tradução nossa), como “uma grande e criteriosa coleção de textos naturais.”⁴

Ao compararmos todas essas paráfrases usadas para estabelecer o conceito de *córpus*, percebemos que há palavras-chave que remetem a aspectos importantes que devem ser levados em consideração, no processo de compilação das amostras, por exemplo, “criteriosa” destaca uma série de cuidados relacionados à representatividade e à finalidade da pesquisa. Ao atentarmos para “eletrônicos”, percebemos a necessidade de que os textos estejam em um formato que pode ser processado por computadores. A palavra “uso”, relacionada à linguagem, denota que os textos precisam servir para a comunicação e interação entre as pessoas em seu cotidiano, isto é, não podem ser artificiais, criados apenas para fins de análises linguísticas. Logo, é evidente que um *córpus* deve ser uma amostra de linguagem natural, que segundo Berber Sardinha (2000, p. 336), “por textos naturais entende-se ‘autênticos’, isto é, aqueles que existem na linguagem e que não foram criados com o propósito de figurarem no *córpus*. Além disso, amplia-se a ideia de ‘natural’ para incluir somente aqueles textos produzidos por humanos.”

Considerando todos os critérios que devem ser seguidos para a formação de *corpora*, é visível que sua compilação muitas vezes pode ser uma tarefa árdua e os inúmeros propósitos científicos a que servem deixam claro que há inúmeros tipos de *corpora*, pois eles podem ser classificados como gerais, de especificidade, monitores, monolíngues, multilíngues, paralelos, dinâmicos, estáticos, de aprendizes, pedagógicos e oportunistas. Um quadro de classificação, considerando apenas o critério da seleção das amostras, pode ser exemplificado, segundo Beber Sardinha, da seguinte maneira:

Quadro 1–Principais tipos de *corpora* de acordo com o critério da seleção das amostras

Tipo	Conceito
Amostragem	É composto por porções de textos 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 <i>corpora</i> de amostragem.
Dinâmico	O crescimento e diminuição são permitidos, qualifica o <i>corpus</i> monitor.
Estático	Oposto de dinâmico, ele caracteriza o <i>corpus</i> de amostragem.
Equilibrado	Os componentes (gêneros, textos etc.) são distribuídos em quantidades semelhantes (por exemplo, mesmo número de textos por gênero).

Fonte: Berber Sardinha (2004, p. 20).

⁴ A *corpus* is a large and principled collection of natural texts.

Tomando como exemplo o *British National Corpus*, podemos perceber que um único *cópus* pode se adequar a várias tipologias de acordo com os critérios de sua compilação, assim, ao analisar certos aspectos de sua estrutura, percebe-se que o BNC é um *cópus* monolíngue, geral, estático/sincrônico. Monolíngue, por ter sido planejado para representar apenas a variedade britânica da língua inglesa. Geral, pois, de acordo com Sarmiento (2010, p. 91), contém “muitos tipos de textos. Pode incluir linguagem escrita, falada ou ambas; textos produzidos em um país ou vários.” E, segundo Oliveira (2009, p. 58), *corpora* gerais “visam representar a língua de forma ampla e servir de base para pesquisas variadas; eles caracterizam-se pela sua variedade em relação aos gêneros discursivos que incluem a variedade de registros, assuntos e autores.” E, como não pode ser ampliado, ele se adequa a outro conceito de *corpora*, que de acordo com Viana (2010, p. 29), “a maior parte dos *corpora* existentes, no entanto, corresponde a coleções em que a inclusão ou exclusão de textos não é permitida. Esses são denominados de *corpora* estáticos.” Também é um *cópus* de amostragem, pois, de acordo com Berber Sardinha (2000, p. 340), é “composto por porções de textos ou de variedades textuais, planejado para ser uma amostra finita da linguagem como um todo.”

Tomando como parâmetro o COCA, amostra que compõe o material desse trabalho, é um *cópus* monolíngue planejado para ser representativo da língua inglesa em sua variedade norte-americana e se destaca por ser atualizado anualmente, mantendo seu equilíbrio, portanto se trata de um *cópus* dinâmico, sua base de dados conta com uma estrutura de amostragem que permite que seja capaz de fornecer dados para pesquisas sincrônicas e diacrônicas. É um *cópus* monitor, que segundo Sarmiento é:

Planejado para verificar mudanças atuais em uma língua. Esse tipo de *cópus* é alimentado anualmente, mensalmente ou até diariamente, aumentando de tamanho rapidamente. Entretanto, a proporção de tipos de textos mantém-se constante, de forma que cada período de tempo possa ser comparado com o anterior. (SARMENTO, 2010, p. 91-92).

Sua capacidade de permitir a comparação de amostras de determinado período com períodos anteriores é fator decisivo para a realização dessa pesquisa, pois entre seus propósitos estão complementar e atualizar os dados da pesquisa realizada por Liu (2011).

Além dos tipos de *corpora* já explicados há outros tipos que também se destacam por prestarem contribuição significativa para as pesquisas em LC. O quadro 02 traz alguns deles juntamente com as respectivas definições.

Quadro 2 - Outros tipos de corpora

Tipo	Conceito
Falado	Composto de porções de fala transcritas.
Escrito	Composto de textos escritos, impressos ou não.
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.
Especializado	Os textos são de tipos específicos (gêneros ou registros definidos).
Regional ou dialetal	Os textos são provenientes de uma ou mais variedades sociolinguísticas específicas.
Multilíngue	Inclui idiomas diferentes.
De aprendiz	Os autores dos textos não são falantes nativos.
De língua nativa	Os autores são falantes nativos.
Paralelo	Os textos são comparáveis (por exemplo, original e tradução)
Alinhado	As traduções aparecem abaixo de cada linha do original
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, p. 20-21).

Além dos tipos mencionados, que seguem estruturas de amostragem previamente planejadas, para serem representativas de determinada variedade linguística, criteriosamente compilados para alcançarem abrangência e equilíbrio, servindo para finalidades de pesquisa adequadamente estabelecidas através do processo de sua compilação; podem ser encontrados em pesquisas da Linguística de *Córpus*, de acordo com McEnery e Hardie, os *corpora* oportunistas, que:

Não têm pretensão de aderir a rigorosas estruturas de amostragem, nem aspiram lidar com questões de distorção pela coleta de um *córpus* cada vez maior de dados, como monitor *corpora* podem. Em vez disso, eles representam nada mais nada menos do que os dados que foram possíveis reunir para uma tarefa específica.⁵ (MCENERY; HARDIE, 2011, p. 11).

Depreende-se então que, apesar de não se ajustarem aos modelos de coleta de amostragem associados a práticas tidas como ideais para LC, os *corpora* oportunistas desempenham um papel valioso para as pesquisas linguísticas, uma vez que compreendem amostras limitadas, mas raras de línguas que já caíram em desuso. Além da escolha pelo tipo

⁵*These corpora make no pretension to adhere to a rigorous sampling frame, nor do they aspire to deal with issues of skew by the collection of an ever-larger body of data, as monitor corpora may. Rather, they represent nothing more nor less than the data that it was possible to gather for a specific task.*

de *córpus* que melhor se adéqua à determinada pesquisa, é importante adotar cuidados no sentido de se evitar distorções nos dados.

Nesse contexto, destacamos a questão da representatividade e se a pesquisa demanda uma amostra anotada ou crua (não anotada). Questões referentes à representatividade e à anotação estão detalhadas na sequência.

Um conceito vital relacionado à compilação de *corpora*, diz respeito à representatividade e é muito debatido na LC, pois afirmar que uma amostra representa uma determinada língua ou variedade linguística denota cautela, pois capturar a linguagem em sua completude ainda compreende uma tarefa utópica. Nesse sentido, Berber Sardinha (2000, p. 342) afirma que “o *córpus* é uma amostra de uma população cuja dimensão não se conhece a linguagem como um todo. Desse modo, não se pode estabelecer qual seria o tamanho ideal da amostra para que ela represente essa população.” Assim, justifica-se o emparelhamento frequente de “coleção” e “criteriosa” ao definir o que é um *córpus*.

Relacionar representatividade à extensão da amostra parece ser o caminho mais sensato, pois, se o léxico de uma língua é composto tanto por palavras de alta frequência quanto das menos frequentes, sua amostra deve capturar palavras dos dois grupos. Logo, uma forma de superar a questão da representatividade é buscar uma amostra com extensão suficiente para capturar o maior número de palavras raras, pois como afirma Berber Sardinha (2000, p. 342), “quanto maior a quantidade de palavras, mais probabilidade há de palavras de baixa frequência aparecerem.” Então, segundo o mesmo autor, “quando se diz que um *córpus* deva ser representativo, entende-se representatividade em termos da extensão do *córpus*, isto é, de um número determinado de palavras e de textos.” (BERBER SARDINHA, 2000, p. 343)

Outro aspecto relacionado à compilação de *corpora* é a anotação. Segundo Berber Sardinha (2012, p. 328), consiste em “adicionar informação aos arquivos coletados.” Essa adição de informação aos *corpora* pode ser muito útil dependendo da natureza da investigação que será realizada, então, percebe-se que não compreende uma prescrição, logo, ao analisar a literatura da LC nota-se que há coletâneas anotadas e outras não anotadas, ou cruas. Há dois tipos comuns de anotação de *corpora*, o *tagging* e o *parsing*. Este, segundo Berber Sardinha (2012, p. 330), é um “tipo de anotação que adiciona aos textos a análise sintática de cada frase dos textos.”, enquanto aquele é uma etiquetagem gramatical, isto é, morfossintática, em que as unidades linguísticas recebem uma identificação referente à sua classe gramatical (BERBER SARDINHA, 2012, p. 329-330). Há três formas de proceder a anotação gramatical, uma realizada apenas pela máquina, a segunda apenas pelo ser humano, e a terceira conta com a união da máquina e do homem, nenhum dos três caminhos alcança o

nível total de acerto, devido às limitações da máquina e do homem. Nesse sentido, de acordo com Aluísio e Almeida:

A anotação linguística pode ser em qualquer nível que se queira, isto é, nos níveis morfossintáticos, sintático, semântico e discursivo, etc., sendo inserida de três formas: manualmente (por linguistas), automaticamente (por ferramenta de processamento de língua natural – PLN) ou semi-automaticamente (correção manual da saída de outras ferramentas). (ALUÍSIO; ALMEIDA, 2006, p. 161).

No que se refere à etiquetagem, duas correntes de pesquisa destacam, a saber, a denominada *corpus-driven* ou dirigida pelo cópuz, em que a etiquetagem é desprezada; e a conhecida como *corpus-based* ou baseada em cópuz, nessa corrente, a etiquetagem é essencial. Há muito mais conceitos envolvendo esses dois modos de investigação linguística em LC, eles serão discutidos nos próximos parágrafos que se destinam à conceituação da LC.

2.1.2.2 Linguística de Cópuz

Com relação à LC, parafraseando Berber Sardinha (2009, p. 7), a LC pode ser entendida como um campo que se dedica à criação e análise de cópuz, que vem mudando a maneira como se investiga a linguagem, nos seus mais diversos níveis, colocando à disposição do linguista quantidades de dados antes inacessíveis. Os linguistas podem lidar com milhões de palavras em segundos e as análises não estão limitadas apenas a aspectos quantitativos, uma vez que os fenômenos capturados através dos cópuz permitem análises qualitativas que enriquecem substancialmente o arcabouço da Linguística.

Para Flowerdew (2009, p. 393, tradução nossa), a LC “é usualmente associada como uma abordagem fraseológica para análises, a qual toma uma visão de linguagem sintagmática, como oposta a uma puramente paradigmática.”⁶

De acordo com a concepção de Biber e Rappen (2015, p. 1, tradução nossa), A LC pode ser entendida como “[...] uma abordagem de pesquisa que facilita investigações empíricas da variação e uso da língua, resultando em descobertas que têm generalizabilidade e validade muito maiores que diferentemente seria viável.”⁷

De acordo com McEnery et al. (2019, p. 74), a LC é um modelo quantitativo baseado em uma tradição de análise da linguagem. Ela usa grandes quantidades de dados

⁶ *Corpus linguistics is usually associated with a phraseological approach to analysis, which takes a syntagmatic, as opposed to a purely paradigmatic, view of language.*

⁷ *Corpus linguistics is a research approach that facilitates empirical investigations of language variation and use, resulting in research findings that have much greater generalizability and validity than would otherwise be feasible*

observacionais compilados em conjuntos de dados, chamados *corpora*, para prover evidências sobre o uso da língua por ambos falantes de L1 e L2.

Na perspectiva de McEnery e Wilson (2001, p. 1, tradução nossa), “é o estudo da língua baseado em exemplos reais de uso da linguagem.”⁸

Segundo Kennedy (1998, p. 1, tradução nossa), “é não um fim em si própria, mas é uma fonte de evidência para melhorar descrições de estruturas e usos da língua e para várias aplicações, incluindo o processamento da linguagem natural pela máquina e entendendo como aprender e ensinar a língua.”⁹

Por meio da análise dos conceitos mencionados, fica clara a natureza empírica dos estudos da LC, destacam a possibilidade de generalizações, uma vez que os *corpora* representam a melhor alternativa criada até agora com a finalidade de capturar amostras linguísticas, realçam a ideia de linearidade, que nos arcabouços teóricos da LC, evidenciam a natureza probabilística que envolve as associações de palavras; e a preocupação com a representatividade da amostra.

Para processar amostras gigantescas da linguagem, podem ser utilizadas ferramentas de busca presentes na própria estrutura de alguns *corpora* desenvolvidos para serem acessados e processados no ambiente virtual da internet, no caso de *corpora* “caseiros”, isto é, compilados para pesquisas específicas pontuais, o pesquisador precisará utilizar ferramentas desenvolvidas para esse propósito, mas algumas precisam ser compradas, como é o caso do *WordSmith Tools*, que de acordo com Berber Sardinha (2009, p. 8), “é um conjunto de programas integrados (‘suíte’) destinado à análise linguística. [...] permite fazer análises baseadas na frequência e na co-ocorrência de palavras em *corpora*.” Para suprir as necessidades dos linguístas de cópuz, é necessário que tais ferramentas possam gerar listas de frequência, apresentam todas as palavras do cópuz em ordem crescente ou decrescente, respeitando a frequência ou em ordem alfabética; concordâncias, ou melhor, “listagens de palavras de um cópuz, ladeadas por porções dos textos em que ocorrem, feitas por computador.” (BERBER SARDINHA, 2010, p. 294); e listas de palavras-chave, que de acordo com Berber-Sardinha (2009, p. 193), “são aquelas cujas frequências são diferentes, de modo estatisticamente significativo, das frequências utilizadas em um cópuz de referência.” Esses recursos permitem que os linguístas explorem a frequência e a co-ocorrência das palavras, que de acordo com Berber Sardinha (2010, p. 296), são “fontes primordiais de dados

⁸ [...]the study of language based on examples of real life language use.

⁹ Corpus linguistics is not an end in itself but is one source of evidence for improving descriptions of structure and use of languages and for various applications, including the processing of the natural language by machine and understanding how to learn or teach a language.

da LC, são elementos vitais do sistema linguístico.” Pois, ainda tomando emprestado as palavras do autor, que ao destacar o comportamento linguístico dos usuários de uma língua, afirma que “a maioria dos falantes nativos tende a seguir padrões léxico-gramáticos semelhantes, na maioria das vezes, o que aparece nas análises de *corpora* como sendo os padrões mais típicos de uma variedade linguística em particular (BERBER SARDINHA, 2010, p. 297).

Portanto, as associações, não são casuais, mas sim, refletem usos probabilísticos influenciados pelo contexto. Ao estudar as relações que interligam léxico, contexto e registro, é possível dizer se uma determinada unidade lexical ou um determinado emparelhamento se destaca na oralidade ou na escrita, e até mesmo, em contextos de uso formal ou informal.

As pesquisas realizadas por Sinclair focavam no léxico e na colocação, usavam *corpora* crus, eram chamadas de *corpus-driven*, isto é, conduzidas pelo cópuz, eram indutivas e partiam do interior do cópuz. Diferentemente, há uma corrente, mais comum, que realiza pesquisas classificadas como *corpus-based*, ou melhor, baseadas no cópuz, elas, ao contrário estão focadas na variação, utilizam *corpora* anotados e privilegiam a gramática, atualmente, são bem ilustradas pelas pesquisas realizadas por Douglas Biber.

Um traço linguístico que merece destaque por sua contribuição nos estudos da LC é a colocação, que segundo McEnery e Wilson (2001, p. 85), é entendida como “os padrões de co-ocorrência característicos das palavras¹⁰”, pois os estudos da colocação possibilitaram a descoberta de unidades fraseológicas mais frequentes, que ao serem disponibilizadas, principalmente em dicionários, na área da lexicografia, proporcionaram a maior disseminação de informações que além de evidenciarem que em nossa mente, o processo de formação linear na linguagem ocorre de forma mais complexa do que a simples justaposição de unidades isoladas da língua, ainda fortaleceu o argumento de que, para se alcançar a tão almejada fluência em uma língua, é preciso ir além da memorização de formas e sentidos isolados, pois as palavras não apenas ocorrem em um texto, mas acima de tudo, elas co-ocorrem.

Sinclair (1991), ao considerar a colocação e o processo de significação e interpretação de textos, afirma que há dois princípios básicos, o primeiro, o princípio da escolha aberta, é a maneira comum de se enxergar a linguagem. Segundo as palavras do próprio autor (SINCLAIR, 1991, p. 109, tradução nossa), “é frequentemente chamado de ‘*slot-and-filler*’,

¹⁰ *Collocations – the characteristic co-occurrence patterns of words.*

visualizando o texto como uma série de lacunas que precisam ser preenchidas com um léxico que atenda a restrições locais¹¹”.

Nessa perspectiva, lidamos com as palavras individualmente; o segundo, o princípio do idioma, “é que um usuário da língua tem a sua disposição um grande número de frases semi-pré-construídas que constituem escolhas individuais, mesmo que elas possam parecer analisáveis em segmentos¹²” (SINCLAIR, 1991, p. 110, tradução nossa).

O segundo princípio revela o conceito de colocação, uma vez que, compreende-se que as escolhas podem ser representadas não apenas por uma única palavra, mas sim, por grupos de palavras, como os *phrasal verbs*. O conhecimento dos colocados, isto é, a consciência de como as palavras se associam, é um importante aporte para a desenvolvimento da fluência. Os estudos com corpora representam uma fonte confiável para desvendar quais palavras são as mais frequentes e com que probabilidade elas co-ocorrem.

Ainda argumentando sobre os processos de construção de sentidos, Sinclair (2004, p. 134, tradução nossa) afirma que “situações frequentemente surgem no texto onde o sentido preciso de uma palavra ou frase é determinado mais pelo ambiente verbal do que por parâmetros de uma unidade lexical.”¹³

O autor deixa claro que, para extrair os sentidos de determinados textos, não basta apenas decodificar as palavras individualmente, pois há contextos em que as palavras ganham sentido metafórico. Faz-se necessário, então, lançar mão de estratégias de interpretação que levam em consideração os emparelhamentos, ao invés de focar em palavras isoladas.

Com as contribuições da LC, torna-se possível superar a visão simplória de que idiomas são aprendidos por meio de memorização de palavras isoladas, como enfatiza Berber Sardinha:

Normalmente, os aprendizes de língua estrangeira tendem a ver a linguagem como um conjunto de palavras individuais sustentadas por regras gramaticais; desse modo a compreensão e a produção se dão, em muitos casos, com base na junção de palavras individuais. O resultado disso é que, por exemplo, a fluência, um aspecto natural da fala na língua materna, é conseguida somente a duras penas da língua estrangeira. (BERBER SARDINHA, 2009, p. 13).

¹¹ It is often called a 'slot-and-filler' model, envisaging texts as a series of slots which have to be filled from a lexicon which satisfies local restraints.

¹² The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.

¹³ Situations frequently arise in texts where the precise meaning of a word or phrase is determined more by the verbal environment than the parameters of a lexical entry.

Nesse sentido, é importante destacar que os PV, são grupos de palavras com baixo grau de previsibilidade semântica, e seus sentidos não são decifráveis pela mera tradução de suas partes. Para esclarecer melhor o que são PV, buscamos uma definição através das palavras de Moon (2011, p. 45), que ao tratar das conexões vocabulares e “multipalavras”, explica os *phrasal verbs* de forma simplificada, como sendo combinações de verbos e partículas adverbiais ou preposicionais. Sua visão harmoniza-se com Harmer (2015, p. 69) que afirma que *phrasal verbs* são formados pela adição de uma partícula (advérbio ou preposição – ou um advérbio e uma preposição) a um verbo para criar novos significados.

Carter e McCarthy (2006, p. 431) também os definem de modo semelhante ao dizerem que *phrasal verbs* consistem de um verbo lexical e uma partícula. Essas três definições explicam claramente o processo de formação de um *phrasal verb*, mas apenas esse conhecimento pode levar à identificação de estruturas semelhantes, tomadas como *phrasal verbs* de forma equivocada, pois há um grupo de verbos na língua inglesa, os verbos preposicionais, que também são seguidos por preposição, mas não são *phrasal verbs*, uma vez que a relação entre o verbo e a partícula apenas cumpre um papel sintático, não alterando o sentido original do verbo lexical, um exemplo retirado do *corpus* é “*May be you are the wrong guy to go on the raft.*”

Nesse contexto, “go” continua significando “ir” e “on” apenas se associa a ele para indicar a posição “sobre, em cima” no caso em questão “ir em cima da jangada”. Além da relação semântica relacionada aos *phrasal verbs*, vale destacar as variações de ordem sintática, que dificultam a aprendizagem dessas unidades linguísticas e ainda colaboram para que os aprendizes tomem uma posição de relutância e aversão frente aos *phrasal verbs*.

Harmer (2015, p. 69-70) destaca quatro tipos de *phrasal verbs*, a saber, intransitivos, transitivos inseparáveis, transitivos separáveis e, finalmente, transitivos com duas partículas inseparáveis. Com relação aos PV transitivos separáveis, vale lembrar que podem aparecer separados por intervalos formados por várias palavras. Portanto, PV podem ser categorizados entre as palavras do léxico da língua inglesa que mais desafiam os aprendizes, pois, além de todos os aspectos mencionados, ainda é preciso adequar seus usos aos contextos corretos.

Nessa perspectiva, Silva (2012, p. 57), afirma que “[...] conhecer uma palavra envolve saber sua demanda no cotidiano e quais as palavras mais prováveis se associam a ela, como também seus diversos significados.”Essas e outras questões são objeto de discussão da seção 2.1.3 que trata das relações entre LC e Ensino.

2.1.3 LC e Ensino

A LC tem atestado, por meio de inúmeras pesquisas com *corpora*, que as informações extraídas dos padrões linguísticos observáveis em concordâncias, isto é, os colocados, *clusters*, coligações e demais aspectos relacionados à léxico-gramática, juntamente com as descobertas associadas às listas de palavras e suas frequências de uso, compõem um aporte relevante para a consecução de projetos didático-pedagógicos. Nesse sentido, Oliveira afirma que:

As aplicações pedagógicas de estudos do *cópus* podem ir além das descrições linguísticas, tendo impacto direto no planejamento de currículos e nas práticas pedagógicas ligadas ao ensino de línguas. Estas aplicações ilustram a interface entre a Linguística de *Cópus* e a Linguística Aplicada e trazem à tona, por exemplo, pontos mais relevantes e realistas da gramática para o estudo em sala de aula. (OLIVEIRA, 2009, p. 54).

Assim, não é surpresa produtores de materiais didáticos, grandes editoras ao redor do mundo, terem enxergado nos *corpora* e em suas descobertas, a oportunidade de tornarem seus produtos mais rentáveis, uma vez que passaram a refletir os usos da linguagem, ao invés de recorrerem a amostras artificiais da linguagem. Porém, as contribuições da LC para o ensino não se limitam a alimentar páginas de dicionários, gramáticas e livros didáticos com padrões reais de uso de determinada língua.

Nessa direção, Johns e King (1991b apud BERBER SARDINHA, 2010, p. 300) advogam a favor de um modelo de ensino chamado *Data Driven Learning*, ou DDL que corresponde ao “uso, em sala de aula, de concordâncias geradas por computador com a finalidade de fazer com que os alunos explorem as regularidades de padronização da língua alvo, bem como o desenvolvimento de atividades e exercícios baseados em concordâncias.” Essa modalidade de ensino, de acordo com Berber Sardinha (2010, p. 294), é também conhecida como “*classroom concordancing*, pois fundamenta-se no uso das concordâncias [...] como instrumento de ensino.” E segundo o mesmo autor, os materiais desenvolvidos para essa filosofia de ensino são referidos por ele como “centrados na concordância”, visto que ela é a peça central, senão a única, da atividade” (BERBER SARDINHA, 2010, p. 295).

Apesar das dificuldades encontradas nas salas de aula brasileiras para a implantação do ensino de línguas em DDL, pois as escolas do país ainda não contam com internet e aparelhagem adequadas, caso investimentos sejam realizados para suprir essas deficiências, tanto professores quanto alunos podem vir a se beneficiar desse modelo, devido à tomada de consciência acerca da natureza probabilística da linguagem, que segundo afirma, Berber Sardinha (2010, p. 296), “para usar a léxico-gramática com eficiência, é necessário conhecer

as probabilidades daquelas escolhas, isto é, as frequências dos elementos, suas combinatórias e as frequências destes.” Destarte, percebe-se que o contato das partes envolvidas no processo de ensino-aprendizagem com as concordâncias e os *corpora* é fator que pode impulsionar não só a memorização de conceitos relacionados, mas, acima de tudo, promover a real compreensão de vários aspectos da linguagem, tornando o ensino de línguas mais natural e significativo.

A defesa da utilização da pesquisa no ambiente escolar com a finalidade de efetuar uma mudança no papel do aluno, tornando-o mais ativo e autônomo, não é nenhum fato novo na literatura da área. Por exemplo, Demo (2015, p. 9) afirma que “a aula que apenas repassa conhecimento, ou a escola que somente se define como socializadora de conhecimento, não sai do ponto de partida, e, na prática, atrapalha o aluno, porque o deixa como objeto de ensino e instrução”. Na mesma direção, Berber Sardinha, ao argumentar em prol das vantagens da implantação do DDL no ensino de idiomas afirma que:

O sustentáculo principal é o desejo de tornar o aluno um pesquisador, de tal forma que ele viesse a buscar nas concordâncias regularidades no uso autêntico da língua. Com isso, ele encontraria por si mesmo padrões de uso e notaria nesses padrões a resposta para questões importantes do aprendizado de uma língua estrangeira, como o significado de palavras e expressões, o uso de classes gramaticais, além de questões relativas ao texto acadêmico e à cultura da língua estudada. (2010, p. 299).

Dessa forma, pode-se verificar que o modelo de ensino em DDL pode contribuir para o desenvolvimento da autonomia do aprendiz, ao passo que busca soluções para a superação do ensino baseado na cópia, por meio da implementação de um modelo fundamentado na pesquisa e na produção/descoberta de novos saberes. Um paradigma que segundo Berber Sardinha (2010, p. 300), “o professor também, por sua vez, assume outro papel, o de orientador em vez de detentor das respostas. Ele precisa aceitar que os alunos podem descobrir facetas do uso da língua que ele desconhece.” E que, como acrescenta o autor, “o aluno se torne pesquisador e/ou ‘detetive linguístico’, buscando ‘pistas’ nas concordâncias que o levem a descobrir padrões e sentidos nos dados” (BERBER SARDINHA, 2010, p. 300). Essa questão revela-se um espaço real de atuação para a LC, em que os *corpora* podem figurar como verdadeiros protagonistas no ensino de idiomas. Entendemos, que as contribuições dos estudos de *corpora* para o desenvolvimento de materiais didáticos também não pode ser menosprezada, portanto retomaremos a questão.

Ao argumentar sobre a importância do papel dos estudos da LC no ensino de línguas, McEnery e Wilson (2001, p.120) destacam que “muitos desses estudos têm descoberto que

existem diferenças consideráveis entre o que os livros didáticos estão ensinando e como falantes nativos usam a linguagem como evidenciado nos *corpora*.¹⁴”

Também argumentando sobre a questão da autenticidade da linguagem no contexto escolar e a relevância das pesquisas com *corpora* na área, Meunier e Reppen (2015, p. 501) defendem que “descrições linguísticas baseadas em grandes *corpora* abrangentes e descrições de *corpora* especializados contribuíram muito para o conhecimento das características linguísticas do uso da linguagem em diferentes situações e circunstâncias de produção.¹⁵”

De acordo com o estudo de Biber e Rappen (2002), ao tratarem do papel da frequência do ensino de gramática, afirmam que além do destaque dado para textos autênticos no ensino de línguas, no que tange ao ensino de itens gramaticais, é importante levar em consideração três informações, a saber: quais itens devem ser incluídos na lição ou no programa de ensino; em que ordem esses itens devem ser incorporados e quais itens lexicais devem ser privilegiados nesse processo. Suas análises demonstraram equívocos nos livros didáticos, uma vez que não foram baseados em estudos de *corpora* quando planejados e desenvolvidos, logo as decisões referentes aos três aspectos estudados, comprovam uma discrepância entre o conteúdo dos materiais e o que realmente acontece no uso real da língua. Tal distanciamento entre o inglês dos livros e o do dia a dia dos falantes nativos é consequência de decisões baseadas na tradição e na intuição das pessoas envolvidas no processo. Para garantir que os aprendizes tenham contato com estruturas autênticas e significativas da língua, é essencial que os materiais reflitam usos comuns. Além dessas contribuições, a LC também pode atuar e desempenhar papel relevante em pesquisas da área da Aquisição de Segunda Língua (ASL), por meio da compilação e análises de *corpora* de aprendizes. Nesse sentido, McEnery *at al.* (2019, p. 84) afirmam que “*corpora* e métodos que usam *corpus* podem trazer [...] inovação significativa para as áreas em ASL que já têm consideravelmente se beneficiado de abordagens quantitativas, especificamente, baseadas em *corpus*.¹⁶” Por exemplo, podem explorar, com perguntas específicas de investigações em ASL, a linguagem formulada, isto é, os diversos tipos de emparelhamentos, por meio de *softwares* desenvolvidos para estas análises.

Outra possibilidade de inclusão de *corpora* para a produção de material destinado ao ensino de idiomas é sugerida por Berber Sardinha, Delfino e Rampaso (2017, p. 4), pois, uma

¹⁴Most of these studies have found that there exist considerable differences between what text books are teaching and how native speakers use the language as evidenced in the *corpora*.

¹⁵Linguistic descriptions based on both large comprehensive *corpora* and descriptions of specialized *corpora* have greatly contributed to knowledge of the linguistic characteristics of language use across different situations and production circumstances.

¹⁶*Corpora* and *corpus* methods can bring further, significant innovation to the areas in SLA that have already considerably benefitted from quantitative, and specifically *corpus*-based, approaches.

vez que é fácil ter acesso a *corpora* disponibilizados na *web*, ou até mesmo, tanto professores quanto alunos podem compilar uma coletânea própria. Os autores afirmam que o *cópus* CoEL (*Corpus of English Lyrics*) “é uma alternativa para a produção de material didático de Inglês Geral, onde a música torna-se o elemento central da sala de aula, levando tanto o aluno como o professor a trabalharem com a língua de forma prazerosa, utilizando material autêntico”. Portanto, à medida em que os computadores evoluem e a LC se populariza, maiores são as chances de os *corpora* ocuparem espaço no âmbito educacional. Nesse sentido, Chambers (2019, p. 464) ao tratar da lacuna de pesquisas realizadas por professores não linguístas, posiciona-se favoravelmente ao uso dos *corpora* no ensino ao afirmar que:

o aluno que dominou o uso do *software* de análise de *cópus* e que tem acesso a *corpora* relevantes podia decidir independentemente, fora da sala de aula, quais aspectos do uso da linguagem investigar, assim desenvolvendo autonomia, um outro influente conceito em pesquisas de aprendizagem de línguas.¹⁷ (CHAMBERS, 2019, p. 464).

Portanto, na perspectiva dos autores mencionados, é fundamental preencher os materiais didáticos com amostras linguísticas autênticas, garantindo que o aprendiz tenha contato com exemplos que reflitam o uso da linguagem real, do cotidiano dos falantes, isto é, *English language*, baseada em *cópus*. Também frisaram que levar o *cópus* para dentro da sala de aula juntamente com as ferramentas de análise linguística pode promover experiências de aprendizagem que desenvolvem a autonomia do aprendiz, ao passo que se conscientiza de vários aspectos da linguagem.

2.1.4 Variação linguística nos registros

Nesta seção, abordaremos práticas de pesquisas realizadas no âmbito da LC, cujas investigações direcionam suas lentes para os registros linguísticos e suas variações elucidadas por meio da contagem de suas ocorrências no uso da linguagem. Assim, inicialmente apresentaremos os principais conceitos da área e os aspectos dessa abordagem. Na sequência, destacaremos a chamada análise multidimensional, doravante MD, então, por fim, trataremos das relações entre as variações de registro e o ensino da linguagem.

PV são comuns e realizam diversas funções na linguagem, portanto há aspectos relevantes em seus usos na língua inglesa que podem ser evidenciados por meio da utilização de abordagens direcionadas para o estudo das relações léxico-gramaticais que ocorrem em

¹⁷ *The learner who has mastered the use of corpus analysis software and who has access to relevant corpora could decide independently, outside the classroom, what aspects of language use to investigate, thus developing autonomy, another influential concept in language learning research.*

múltiplos registros, uma vez que os traços linguísticos são situacionalmente impulsionados e exercem funções que satisfazem as necessidades comunicativas dos usuários de uma língua.

Como dominar uma língua demanda associar conhecimentos léxico-gramaticais ao uso da linguagem que também é influenciado pela cultura, analisar a variação de registros relacionada aos PV em diversos registros pode revelar traços linguísticos que além de ampliar o arcabouço da área, pode produzir insumos substanciais para a formação de falantes conscientes de características reais da linguagem, aproximando suas escolhas léxico-gramaticais daquelas mais salientes no desempenho linguístico dos falantes nativos.

De acordo com o documento (Brasil, 2019, p. 87) que regulamenta a formação superior, a formação docente deve ser balizada pela BNCC (Base Nacional Comum Curricular). Uma vez que o documento é recente, pode-se dizer que a educação brasileira passa por um processo de transição para se adequar às novas exigências. Tal adequação, então pressupõe o desenvolvimento de um novo currículo para a formação docente.

O texto também ressalta o pleno desenvolvimento do indivíduo, quer e quer “o estabelecimento das pertinentes competências profissionais dos professores”(BRASIL, 2019, p. 87). Um pouco mais adiante, fica evidente, que a formação docente precisa garantir o desenvolvimento de competências previstas na BNCC e também, das aprendizagens fundamentais dos estudantes, levando em conta a evolução não apenas de suas qualidades intelectuais, como também, é preciso tratar de aspectos físicos, culturais, sociais e emocionais, a fim de alcançar uma educação integral. Nesse sentido, a presente pesquisa, ao tratar seu objeto por meio de uma perspectiva de registros, pode gerar conhecimentos necessários para a adequação curricular considerando as novas demandas da educação plena.

Ao privilegiar a perspectiva de registros, em detrimento da perspectiva de gênero (que apesar de seus estudos linguísticos serem significativos, os educadores têm reduzido a área ao simples estudo de características presentes na camada superficial do texto, no modelo vigente de educação tradicional), pode-se contribuir para o estabelecimento de novas práticas docentes no âmbito do ensino das linguagens, oportunizando um paradigma de ensino voltado para o crescimento do ser humano, e que não apenas se contente com a memorização insignificante de conteúdos secundários, mas que realmente produza condições para a compreensão linguística e para sua prática na vida cotidiana.

Ao tratar os PV através da perspectiva de registros, vislumbramos trazer à tona, dados oriundos de investigações empíricas, revelando características inerentes aos seus usos, ao passo que geramos conhecimento atual e autêntico, isto é, que realmente traduz escolhas linguísticas de usuários da linguagem. Caso, esse conhecimento seja anexado aos currículos

da educação básica e superior, contribuirá para a formação de usuários da língua-alvo, ao invés de apenas promover momentos de memorização de conteúdos léxico-gramaticais dispersos dos aspectos sócio-culturais indissociáveis da linguagem tomada como ferramenta de comunicação.

Indubitavelmente, o melhor caminho para expor conceitos-chave presentes nesta seção, precisa ter seu ponto de largada, no termo “registro”, que segundo Biber e Conrad (2009, p.6, tradução nossa), “é uma variedade associada com uma situação particular de uso (incluindo propósitos comunicativos particulares)”¹⁸. Terminologia frequentemente confundida com “gênero”, evidenciando a necessidade de sua desambiguação, nesse sentido, Biber (2010, p.241, tradução nossa) afirma que “na perspectiva do gênero, o foco está sobre as características linguísticas que são usadas para estruturar textos completos.”¹⁹ E, ao definir a perspectiva do registro, o autor afirma que “foca nas características linguísticas dominantes de excertos de textos representativos da variedade.”²⁰ Assim, percebe-se que ao abordar um romance, por exemplo, na perspectiva do gênero, seriam listadas características estruturais convencionais, tais como título, autor, capítulos, enredo, conflitos, enquanto que na perspectiva do registro, seriam elencados traços linguísticos prevalentes em textos narrativos, por exemplo, emprego de pronomes de primeira ou terceira pessoa, verbos no passado etc.

De acordo com Biber e Conrad (2009), a descrição dos registros requer a análise de três aspectos, a saber, primeiro, *os traços linguísticos*, os quais de acordo com os autores são representados pelas “características lexicais e gramaticais típicas dos registros.”²¹(BIBER ;CONRAD, 2009, p. 6, tradução nossa); em segundo lugar, é preciso considerar *os contextos situacionais* de produção dos registros, estes exemplificados pelos autores considerando se foram produzidos na fala ou na escrita, se são interacionais e quais são suas finalidades comunicativas. E, por fim, o processo de descrição dos registros deve considerar *as relações funcionais* presentes tanto nos traços linguísticos como nas condições de produção, que nas palavras dos autores, “os traços linguísticos tendem a ocorrer em um registro porque eles são particularmente adequados para os propósitos e contextos situacionais do registro.”²² (BIBER; CONRAD, 2009, p.6, tradução nossa.) Então, percebe-se que na visão dos registros, os traços

¹⁸ A register is a variety associated with a particular situation of use (including particular communicative purposes).

¹⁹ In the genre perspective, the focus is on the linguistic characteristics that are used to structure complete texts.

²⁰ The register perspective focuses on the pervasive linguistic characteristics of representative text excerpts from the variety.

²¹ Registers are described for their typical lexical and grammatical characteristics: their linguistic features.

²² Linguistic features tend to occur in a register because they are particularly well suited to the purpose and situational context of the register.

linguísticos estudados, são sempre considerados funcionais.

Dentre as pesquisas baseadas em *corpora* destinadas à investigação de variação em registros, destacam-se aquelas que tratam seus objetos de estudo através da análise Multi-Dimensional, que segundo as palavras de Biber e Conrad (2009, p. 223, tradução nossa), “é uma abordagem quantitativa que permite ao pesquisador comparar muitos registros diferentes, com respeito a vários parâmetros linguísticos diferentes – as ‘dimensões’²³”. Estas análises são cruciais para a realização do processo de investigação dos registros, pois abrem caminho para a contagem das frequências que podem revelar quais fatos são comuns e quais são mais raros em determinados registros, uma vez que, segundo Biber (2010, p. 246, tradução nossa), cada dimensão representa uma constelação de traços linguísticos que frequentemente ocorrem nos textos.²⁴”

Considerando a presente pesquisa, vale realçar a relevância que a análise dos registros e suas variações para descrever fatos sincrônicos e diacrônicos inerentes aos usos dos PV em vários registros que compõem a base de dados do COCA, uma amostra da variedade norte-americana da língua inglesa, cuja abrangência pode permitir descrições consistentes dos usos da linguagem, como afirma Biber (2010, p.247, tradução nossa), “estudos multi-dimensionais têm revelado dimensões particulares para as funções comunicativas e prioridades de cada domínio diferente de uso.²⁵” e ainda destaca que “*corpus-driven MD* estudos dos registros do inglês têm revelado ambas, similaridades e notáveis diferenças nas dimensões subjacentes da variação²⁶” (BIBER,2010,p.247,tradução nossa).

No que se refere às potencialidades das pesquisas realizadas com *corpora*, de acordo com Conrad (2011, p. 47, tradução nossa) “técnicas da LC tornam possível descrever padrões em como a linguagem é usada em contextos diferentes – isto é, descrever as escolhas que falantes e escritores tendem a fazer em diferentes condições situacionais e discursivas.²⁷” fato que justifica o percurso metodológico realizado nesta pesquisa, pois, almejamos trazer à tona, aspectos da língua inglesa que possam passar despercebidos no cotidiano dos seus falantes e aprendizes, mas que são marcantes e recorrentes para o

²³ *Multidimensional (MD) analysis is a quantitative approach that allows the researcher to compare many different registers, with respect to several different linguistic parameters – the “dimensions”.*

²⁴ *[...] since each dimension represents a constellation of linguistic features that frequently co-occur in texts.*

²⁵ *MD studies have uncovered dimensions particular to the communicative functions and priorities of each different domain of use.*

²⁶ *Corpus-driven MD studies of English registers have uncovered both surprising similarities and notable differences in the underlying dimensions of variation.*

²⁷ *Corpus Linguistics techniques make it possible to describe patterns in how language is used in different contexts – that is, to describe the choices that speakers and writers tend to make in different situational and discourse conditions.*

exercício da comunicação.

Biber (2010, p. 242), ao destacar pesquisas com a associação de corpora e registros, afirma que há dois tipos de estudos, pois, por um lado, “vários estudos focam nas variantes linguísticas associadas com um traço, usando diferenças de registros como um fator para representar os padrões da variação linguística.²⁸” E, por outro lado, o autor também aponta a presença de estudos que “têm focado no uso de um traço particular da linguagem em um único registro, nesse caso, os objetivos do estudo são descrever ambas, funções discursivas dos traços linguísticos e o próprio registro-alvo.²⁹” O presente trabalho se destaca por propor a análise de um traço linguístico, a saber, os *phrasal verbs*, em vários registros diferentes com o objetivo de descrever os padrões recorrentes de variação.

Biber (2010, p.246, tradução nossa) deixa claro que a “frequência exerce um papel central na análise³⁰” Assim, Biber e Conrad (2009, p. 224, tradução nossa) destacam que “a abordagem multi-dimensional foi desenvolvida para analisar os padrões linguísticos de co-ocorrência associados com variação de registro em termos empíricos/quantitativos.³¹” Portanto, ao lançar mão de tais recursos metodológicos, uma pesquisa pode contribuir massivamente para a produção de dados coerentes relacionados aos usos de traços linguísticos em registros distintos.

Ao associar os recursos da LC e o estudo das variações de registro, por meio de análises quantitativas e qualitativas é possível, não apenas descrever traços ligados diretamente à língua, como também crenças e atitudes dos falantes que optam por uma determinada estrutura léxico-gramatical para se adequar às normas institucionais, e como afirma Conrad (2011, p.48), o “estudo baseado em *cópus* investiga escolhas comuns e incomuns em circunstâncias particulares.³²” Logo, ao tratar os registros considerando traços linguísticos e seus contextos de produção ligados às pressões funcionais que impulsionam seus usos, pode-se revelar facetas importantes da linguagem por meio de análises empíricas.

Biber (1993, p. 222, tradução nossa) destaca que “a análise identifica uma distinção

²⁸ *Several studies focus on the linguistic variants associated with a feature, using register differences as one factor to account for the patterns of linguistic variation.*

²⁹ *However, there are an even larger number of studies that have focused on the use of a particular linguistic feature in a single register; in this case the goals of the study are to describe both the discourse functions of the linguistic feature and the target register itself.*

³⁰ *Frequency plays a central role in the analysis.*

³¹ *The MD approach was developed to analyze the linguistic co-occurrence patterns associated with register variation in empirical/quantitative terms.*

³² *Corpus-based work investigates common and uncommon choices in particular circumstances.*

fundamental entre as complexidades do discurso de registros escritos e falados.³³ Uma vez que a presente pesquisa se propôs a demonstrar características dos PV em registros diferentes, evidenciando seus usos em contextos de fala e escrita, comprovando por meio de dados linguísticos reais que seus usos são mais recorrentes em contextos informais, tratar nosso objeto de pesquisa, lançando mão de uma abordagem focada em registros, proporcionou a descoberta de aspectos que se mostraram preponderantes em determinados contextos e extremamente raros em outros.

Considerando os estudos de Biber (1993, p. 228, tradução nossa), é possível dizer que a análise multidimensional revela “padrões sistemáticos de variação entre registros; que estes padrões podem ser analisados em termos de dimensões subjacentes de variação; e que é necessário reconhecer a existência de um espaço multidimensional para capturar todas as relações entre registros.³⁴” Nesse sentido, o autor lista 5 dimensões que sintetizam “um conjunto distinto de traços linguísticos co-ocorrentes; cada um define um conjunto diferente de similaridades e diferenças entre registros falados e escritos; e cada um tem uma base funcional diferente³⁵” (BIBER, 1993, p. 229, tradução nossa). Pode-se, portanto identificar se determinado PV, por exemplo, é mais utilizado para fins comunicativos ligados a contextos informacionais ou interacionais, narrativos, argumentativos e mais ou menos formais.

Os resultados obtidos nesta pesquisa podem contribuir para a mudança de paradigma não apenas no campo científico (considerando pesquisas baseadas na intuição ou que menosprezem os registros), mas acima de tudo, no âmbito educacional que insiste em reduzir a linguagem a um conjunto de regras gramaticais que não retratam a linguagem em sua completude, ao passo que desprezam a linguagem como ela realmente é, para descrevê-la como ela deveria ser, valorizando aspectos artificiais e rejeitando seus traços naturais. Nesse sentido, Liu (2011, p. 679, tradução nossa) afirma que “Aprendizes e professores de inglês podem escolher usar a lista dos 150 PV mais comuns de modos que melhor se adequem aos seus propósitos de aprendizagem.³⁶” Deste modo, depreende-se que dados de estudos que investigam traços linguísticos que se destacam em registros podem contribuir de modo significativo tanto para o sucesso do trabalho de professores quanto para a eficácia da

³³ *The analysis further identifies a fundamental distinction between the discourse complexities of written and spoken registers.*

³⁴ *These studies [MD] have shown that there are systematic patterns of variation among registers; that these patterns can be analyzed in terms of underlying dimensions of variation; and that it's necessary to recognize the existence of a multidimensional space in order to capture the overall relations among registers.*

³⁵ *Each comprises a distinct set of co-occurring linguistic features; each defines a different set of similarities and differences among spoken and written registers; and each distinct functional underpinnings.*

³⁶ *English learners or teachers can elect to use the lists of the 150 most common PV in ways that best meet their learning purposes.*

aprendizagem dos estudantes.

Liu (2011, p.675, tradução nossa) aponta que “os PV são muito mais comuns na ficção e na fala inglesa do que em revistas, jornais, e, especialmente na escrita acadêmica³⁷”, desta afirmação, pode-se inferir que os estudos empíricos baseados em *corpora* produzem informações que corroboram ou rejeitam o senso-comum, neste caso, fica evidente que os PV atuam de forma substancial nos registros tomados comomais informais da comunicação humana.

Ao destacar o campo semântico das lexias, Liu (2011, p. 676) afirma que “a distribuição de diferentes sentidos de um PV polissêmico pode variar significativamente através dos registros³⁸, então, é notório que há um campo aberto para novas pesquisas, pois os PV são muito comuns na comunicação cotidiana dos falantes nativos e muitos deles são polissêmicos, fato que possibilita que atuem em vários contextos para satisfazer inúmeras necessidades comunicativas dos usuários da língua inglesa. Assim, percebe-se que a abordagem focada em registros pode relevar pontos importantes do uso dos PV como acrescenta, o próprio autor, ao afirmar que os resultados “sugerem claramente que a distribuição dos diferentes significados de um PV através dos registros é também informação importante³⁹”(LIU,2011, p.676, tradução nossa).

Como exemplo das possibilidades de se estudar PV pela perspectiva de registros, especialmente, através da modalidade *cross-register*, podemos citar a descoberta de Liu (2011, p. 677, tradução nossa) que evidenciou a presença de *phrasal verbs* que atuam em todos os registros do *corpus*. Como é o caso de *point out*, destacado na pesquisa como um “PV distribuído bastante equitativamente⁴⁰”, apesar de se destacar sutilmente em textos acadêmicos.

Os PV são lexias complexas, realmente, instigantes, pois além de idiomáticos, possuem usos marcantes devido às possibilidades de posicionamento dentro da estrutura sintática e em diversos contextos de uso da linguagem para satisfazer as variadas necessidades comunicativas do ser humano, portanto suas cargas de aprendizagem tendem a ser extremamente densas. Fato que justifica a necessidade de estudos que possibilitem descobertas de traços típicos de uso que facilitem a formação de usuários não-nativos, que

³⁷The PV are much more common in fiction and spoken English than in magazines, newspapers, and, especially, academic writing.

³⁸The distribution of different meanings of a polyssemous PV may vary significantly across registers.

³⁹Suggest clearly that the cross-register distribution of the different meanings of a PV is also important information.

⁴⁰A fairly evenly distributed PV.

carecem de conhecimento linguístico e cultural para realizar escolhas léxico-gramaticais bem ajustadas com os diversos contextos de uso da linguagem.

Biber (2010, p.242) afirma que “*corpora* foram tradicionalmente muito melhor projetados para a análise de registros do que de gêneros⁴¹”, nesse sentido, quando se trata de investigações de PV, tanto a noção de *corpora* quanto a de registros linguísticos são cruciais, pois estudar PV, por si só, já é uma tarefa desafiadora, pois correspondem a um número muito maior do que os 5000 listados no *Longman Phrasal Verbs Dictionary*, além disso, muitos deles são extremamente polissêmicos, então qualquer empreitada para elucidar dados relacionados a PV, certamente, demandará esforço e dedicação do pesquisador. Lidar com PV por meio de uma abordagem de registros, capaz de revelar seus traços linguísticos, mesmo de um número reduzido, em diferentes registros da linguagem se tornou viável e mais eficiente, devido aos avanços ocorridos na LC e principalmente, com o desenvolvimento da análise multidimensional.

A abordagem de registros também se destaca, por se adequar às ferramentas desenvolvidas para o processamento de *corpora*, como afirma Biber (2010, p. 242, tradução nossa) “ferramentas de *software* como concordanciadores foram desenhados para a análise de características linguísticas predominantes e frequentes (traços de registros) ao invés de traços que ocorrem uma ou duas vezes no texto (traços de gênero)⁴²”, considerando, deste modo, que atualmente contamos com *corpora* que já contam com estas ferramentas introduzidas em suas próprias base de dados, como é o caso do COCA, amostra linguística utilizada no corrente estudo, trata-se de uma amostra extensa e abrangente o suficiente para propiciar as condições necessárias para elucidar características dos usos dos PV que podem contribuir para a adequação curricular proposta pelos documentos reguladores da educação brasileira.

Demonstrando que a realização deste trabalho não está apenas ligada a interesses científicos, pois seus propósitos estão devidamente localizados no tempo e no espaço em que seus empreendimentos científicos ocorrem.

Biber (2010, p. 243) aponta o estudo realizado por Tottie (1991) para exemplificar variações de traços linguísticos que são mais salientes em registros orais e outros que se destacam nos registros da escrita, por exemplo, as formas sintéticas da negação predominam na escrita, enquanto as formas analíticas se mostraram mais proeminentes nos registros da oralidade. Os resultados apresentados evidenciam que recorremos a princípios e recursos

⁴¹*Corpora have traditionally been much better desined for the analysis of register than genre.*

⁴²*Software tools like concordancers have been desined for the analysis of pervasive and frequent linguistic characteristics (register features) rather than features that occur only once or twice in a text (genre features).*

metodológicos coerentes para se lidar com PV afim de elucidar traços linguísticos relacionados aos seus usos na variedade norte-americana da língua inglesa, pois como afirma Biber e Conrad (2009, p. 9, tradução nossa) “as diferenças linguísticas entre registros podem ser derivadas de diferenças situacionais, porque os traços linguísticos são funcionais.”⁴³

Ao destacar alguns estudos que foram fundamentados na perspectiva de registros, Biber (2010, p. 243, tradução nossa) afirma que “muitos estudos analisam uma única categoria de partes do discurso, documentando os padrões de variação e uso em registros particulares.”⁴⁴ O autor, também aborda os estudos que, de acordo com suas palavras “usam diferenças de registro como um indicador de variação linguística, enquanto outros estudam traços linguísticos no contexto de um único registro”⁴⁵ (BIBER, 2010, p. 243, tradução nossa). Destaca também estudos que demonstram que “os tipos de funções dos *lexical bundles* são muito diferentes entre registros da fala e da escrita”⁴⁶(BIBER, 2010, p. 245, tradução nossa). Diante de todas estas possibilidades elencadas, buscamos identificar padrões de uso de um único traço linguístico, a saber, os PV, mas que representam milhares de lexias complexas, e que acima de tudo, são comuns e atuam em vários registros da língua inglesa. Justificando, desse modo, o recorte sugerido pela lista de 150 PV mais comuns elencado por Liu (2011), além da abordagem *cross-register*.

De acordo com Biber (1993, p.226, tradução nossa), “para muitas palavras não há padrão geral de uso que sustenta a linguagem inteira; ao contrário, diferentes sentidos e padrões colocacionais são fortemente preferidos em diferentes registros.”⁴⁷ É o caso de nosso objeto de estudo, pois os PV, além de muitos possuírem campos semânticos amplos, eles aparecem em variadas estruturas léxico-gramaticais, formando blocos sintáticos heterogêneos. Com relação ao fato de nosso caminho metodológico estar ancorado na análise de vários registros, evidencia nossa preocupação por produzir dados confiáveis, uma vez que ao tratar de estudos limitados a um único registro, Biber (1993, p. 228, tradução nossa) afirma que “um corpus restrito a um único registro capacitaria no melhor uma análise parcial do uso lexical; e se os

⁴³*The linguistic differences among registers can be derived from situational differences, because linguistic features are functional.*

⁴⁴*Several studies analyse a single part-of-speech category, documenting the patterns of variation and use in particular registers.*

⁴⁵*Use register differences as a predictor of linguistic variation, while others study linguistic features in the context of a single register.*

⁴⁶*Corpus studies have shown, however, that the types of functions of lexical bundles are very different among spoken and written registers.*

⁴⁷*For many words there is no general pattern of use that holds across the whole language; rather, different word senses and collocational patterns are strongly preferred in different registers.*

resultados fossem generalizados para a linguagem inteira, eles estariam incorretos.⁴⁸” Para realizar seus estudos baseados na abordagem de registros, Biber (2010, p.246, tradução nossa) emprega a análise multidimensional definida pelo autor como “uma *corpus-driven* abordagem metodológica que identifica os padrões linguísticos de co-ocorrência frequentes em uma linguagem, apoiando-se em análises indutivas empíricas/quantitativas.⁴⁹”

Nesta análise, as variações são agrupadas de acordo com as características em cinco dimensões que segundo Biber (1993, p.231, tradução nossa), “representamos padrões linguísticos de co-ocorrência através dos textos, desconsiderando suas categorias de registros.⁵⁰” O autor ainda acrescenta que “os registros são definidos em termos de suas características situacionais, e eles podem ser analisados em muitos diferentes níveis de especificidade⁵¹” (BIBER, 1993, p.233, tradução nossa). As dimensões utilizadas na análise multidimensional estão demonstradas no quadro a seguir:

Quadro3 –Dimensões da análise Multidimensional	
Dimensões da análise Multidimensional	
Dimensão1	Informational versus Involved Production
Dimensão2	Narrative versus Non narrative Concerns
Dimensão3	Elaborated versus Situation-Dependent Reference
Dimensão4	Overt Expression of Persuasion
Dimensão5	Abstract versus Non abstract Style

Fontes: Biber (1993, p. 229)

Ao tratar da análise multidimensional, Biber (1993, p. 221, tradução nossa) afirma que os traços linguísticos individuais são distribuídos diferentemente através dos registros, [...] o mesmo (ou semelhante) traço linguístico pode ter diferentes funções em diferentes registros.⁵²” A análise multidimensional dos PV pode revelar aspectos linguísticos relevantes para o conhecimento dos usos de determinados PV, gerando dados baseados na descrição de fragmentos legítimos de seus usos na língua inglesa. Logo, os resultados obtidos nestes tipos de pesquisa podem servir de aporte para a adequação do currículo de ensino da língua inglesa no Brasil.

O presente trabalho se destaca, pois lida com os PV em uma perspectiva de registro,

⁴⁸A corpus restricted to only one register would enable at best a partial analysis of lexical use; and if the results were generalized to the entire language, they would be incorrect.

⁴⁹Multi-Dimensional (MD) analysis: a corpus-driven methodological approach that identifies the frequent linguistic co-occurrence patterns in a language, relying on inductive empirical/quantitative analysis.

⁵⁰Represent the linguistic co-occurrence patterns across texts, regardless of their register category.

⁵¹Registers are defined in terms of their situational characteristics, and they can be analyzed at many different levels of specificity.

⁵²Individual linguistic features are distributed differently across registers, [...] the same (or similar) linguistic features can have different functions in different registers.

“[...] examinando os traços lexicais e gramaticais predominantes na variedade⁵³” (CONRAD 2011, p.57, tradução nossa). Ao recorrer a esta metodologia, tornou-se possível identificar traços mais comuns em determinados registros e mais raros em outros. Tais variações acontecem porque, de acordo com Conrad (2011, 57) os traços linguísticos “exercem funções importantes para aquele registro – funções modeladas pelo contexto situacional.⁵⁴”

O foco em um traço linguístico, à luz de seus usos em diversos registros, também revela a pujança da pesquisa realizada. Tal empreendimento visa alcançar elevado nível de credibilidade na descrição dos resultados obtidos, pois, no Brasil, a educação está passando por um processo de transição que requer a adequação do currículo de ensino, que no âmbito das linguagens pressupõe que aprendizagens proporcionem o desenvolvimento cultural e social do estudante, logo este currículo precisa ser preenchido com conhecimentos acerca do uso da linguagem.

2.1.4.1 Variação Linguística e Ensino

De acordo com Conrad (2011, p. 59, tradução nossa), “contanto que professores ou professores em formação são introduzidos às técnicas em uma sequência lógica que gradualmente construa suas habilidades e autonomia, eles podem ser introduzidos à Linguística de Córpus em qualquer nível.⁵⁵” A autora destaca que o contato com a LC pode torná-los mais conscientes do funcionamento da linguagem, ao passo que investigam e identificam padrões de uso, logo se percebe que a LC pode contribuir para a formação de professores-pesquisadores, que não só transmitem receitas prontas e tidas erroneamente como verdades absolutas, mas que, fundamentalmente sejam capazes de refletir sobre a linguagem e construir conhecimento atual, deixando de ensinar estereótipos que apenas distanciam os aprendizes da compreensão do curso natural da linguagem e tornam a aula de língua estrangeira tão tediosa e sem sentido.

De acordo com Conrad (2011, p.47, tradução nossa), “os professores ensinavam estruturas gramaticais exatas e corrigiam as estruturas agramaticais que seus alunos produziam.⁵⁶” Depreende-se que a tradição do ensino da linguagem baseava-se na reflexão acerca das regras de como a linguagem deveria ser, em detrimento de como a comunicação

⁵³[...] *examining the pervasive lexical and grammatical features in the variety.*

⁵⁴*Play important functions for that register – functions shaped by the situational context.*

⁵⁵*As long as teachers or teachers or teachers-in-training are introduced to techniques in a logical sequence that gradually builds their skills and autonomy, Corpus Linguistics they can be introduced at any level.*

⁵⁶*Language teachers taught accurate grammatical structures and corrected ungrammatical structures that their students produced.*

realmente acontece, uma perspectiva que segundo as palavras da própria autora, “não pode representar a variação que existe na linguagem, como ela é usada na comunicação ocorrendo naturalmente⁵⁷” (CONRAD, 2011, p. 47), e que jamais encerrará em seu estudo todas as facetas relacionadas à linguagem como ela realmente é. Trata-se de uma visão da linguagem tão limitada que reduz seu objeto a descrições baseadas na precisão gramatical em detrimento de traços essenciais para a descrição linguística, a saber, a adequação aos diversos contextos, situações e finalidades do uso da língua inglesa. Nesse sentido, Biber (1993, p. 220, tradução nossa) afirma que “uma descrição completa da linguagem frequentemente implica uma análise composta de traços como eles atuam em vários registros.⁵⁸” Evidenciando assim, a necessidade de se introduzir no âmbito da educação conhecimento linguístico resultante de pesquisas voltadas para a descrição da variação de registros.

Considerando o ensino de língua inglesa e os PV, Liu (2011, p. 677, tradução nossa) afirma que “claramente [...] a lista lematizada dos 150 PV é uma fonte útil para a aprendizagem dos PV mais comuns, em geral.⁵⁹” O autor ainda acrescenta que “os aprendizes de inglês podem ainda precisar buscar por informações semânticas ou de uso quando aprendendo estes PV⁶⁰” (LIU, 2011, p.677, tradução nossa). Demonstrando que seu empreendimento realizado para produzir informação acerca dos PV em uma perspectiva *cross-registers* pode contribuir para o ensino de língua inglesa.

Ao argumentar acerca de PV que ocorrem em apenas um registro, Liu (2011, p. 676) afirma que “eles são de fato muito importantes para o inglês para propósitos específicos (ESP) aprendizes, que, por causa de seus propósitos específicos de estudos, devem focar em registros em que estes PV aparecem mais frequentemente.⁶¹” Por exemplo, se o aprendiz vai utilizar a língua inglesa para fins acadêmicos, é importante que ele direcione seus estudos para o domínio dos PV que são mais recorrentes em textos da academia. Mais uma vez, evidenciando a utilidade da inserção de resultados de pesquisas fundamentadas na abordagem de registros no âmbito da educação.

É imprescindível estabelecer vínculos entre teoria e prática no espaço de formação docente, nesse sentido, em consonância com as competências elencadas pelo documento Brasil (2019, p.87) os resultados do presente trabalho contribuem para o domínio dos

⁵⁷*It cannot account for variation that exists in language as it is used in naturally-occurring communication.*

⁵⁸*A complete description of the language often entails a composite analysis of features as they function in various registers.*

⁵⁹*Clearly, [...] the lemmatized list of the 150 PV is useful source for learning the most common PV in general.*

⁶⁰*English learners may still need to seek for a semantic or usage information of the PV when learning these PV.*

⁶¹*They are actually very important for English for specific purposes (ESP) learners, who, because of their specific purpose of study, must focus on the register(s) in which these PV appear most frequently.*

conhecimentos acerca dos usos dos PV, para o planejamento das práticas de ensino e, por meio da formação de professores-pesquisadores, ao desafiar os futuros profissionais a refletirem sobre os fenômenos da linguagem pode-se torná-los mais comprometidos com o próprio desenvolvimento profissional.

De acordo com Conrad (2011, p. 58), o estudo das variações de registro deixa claro que “há diferenças na frequência e uso de gramática e vocabulário que são importantes para considerar no ensino de língua.⁶²” Portanto, professores e aprendizes de língua inglesa podem se beneficiar, caso descrições de variação sejam introduzidas na educação básica e superior.

Através de *corpus-based* estudos, em especial, daqueles que tratam da variação de registros considerando a análise multidimensional, pode-se revelar aspectos sincrônicos dos usos da linguagem, uma vez que é possível perceber diferenças de usos entre profissionais de determinada área e estudantes universitários que pretendam atuar em tal área, através da comparação de aspectos presentes em seus textos, sejam escritos ou orais, dependendo apenas da compilação de *corpora* planejados para a realização do estudo.

A compreensão dos fatos linguísticos revelados pelas análises de registros pode contribuir para a formação plena do ser humano, uma vez que o preenche com conhecimento que não está sepultado em livros teóricos, mas que pulsa no cotidiano, pois reflete a comunicação em seu curso natural.

Portanto, a LC possui um arcabouço teórico-metodológico robusto, que evoluiu ao longo das últimas décadas, proporcionando aos estudiosos da linguagem as condições favoráveis para que descrições cada vez mais completas sejam realizadas, revelando traços dos diferentes usos da linguagem, em contextos variados através da perspectiva de registros.

Vale destacar que os estudos alicerçados nas análises multidimensionais ajudam a revelar características que podem estar presentes em diferentes textos, pois estão intrinsecamente ligadas aos contextos de produção e às funcionalidades da linguagem. Logo, abordar os PV através da perspectiva de registros e suas variações pode contribuir para a geração de conhecimento relevante para se alcançar as demandas do ensino de língua inglesa, especialmente, no Brasil, onde os documentos reguladores da educação demonstram preocupação como desenvolvimento pleno do ser humano.

Nesse sentido, destacamos novamente que dominar uma língua pressupõe não apenas conhecer as palavras, as regras sintáticas que regem o modo como elas se associam e os sentidos que transmitem, mas, acima de tudo é preciso saber empregar as palavras

⁶²There are differences in the frequency and use of Grammar and vocabulary that are important to consider in language teaching.

adequadamente para se comunicar de modo eficiente. Logo, é preciso conhecer traços linguísticos relacionados ao uso lexical, suas tendências, preferências, frequências, ocorrências e co-ocorrências para que os aprendizes consigam utilizar a língua inglesa de modo coerente, adequando suas falas aos diversos contextos de uso.

2.2 Interculturalidade

Nesta seção, apresentamos os conceitos e pressupostos teóricos de uma área de estudos da linguística denominada Interculturalidade. A fim de promover uma visão que, ao mesmo tempo, fosse clara e concisa, porém sem deixar de lado conhecimentos importantes para a compreensão da área, primeiramente, tratamos do conceito de cultura – vários autores abordam o tema de modo dicotomizado, isto é, seus elementos são tratados por meio de oposições como visível x invisível, universal x específico, humanística x social. Na sequência, a interculturalidade e suas relações com a linguagem são elucidadas, destacamos que tais ligações não estão limitadas apenas às relações entre pessoas falantes de línguas-culturas diferentes. Para facilitar o entendimento das relações entre a língua e a cultura, apresentamos um quadro do lexicólogo australiano Paul Nation, cuja visão acerca do domínio de unidades linguísticas requer o entendimento da forma, dos sentidos e dos usos das palavras. Finalmente, na última seção, tratamos da interculturalidade e do ensino, destacando a influência dos componentes sócio-culturais para a compreensão da linguagem, salientando a relevância do conhecimento das variações nos diferentes registros linguísticos para o desenvolvimento da competência intercultural.

Assim, como pretendemos lidar com os *phrasal verbs* da língua inglesa em uma perspectiva que contemple seus usos, é fundamental lançar mão dos pressupostos teóricos que capacitem o estudioso da linguagem a trazer à tona, fenômenos ligados aos usos apropriados de diferentes estruturas linguísticas em registros e situações diversificadas e variáveis, pois como afirma Hinkel (2014, p. 400, tradução nossa) “o que faz uma expressão ou ato de fala situacionalmente apropriados não é tanto a forma linguística ou a amplitude do repertório linguístico do falante da L2, mas as variáveis socioculturais, que são raramente abordadas em instruções explícitas.”⁶³ Logo, ao propor a junção dos estudos da Interculturalidade e da LC, por meio das lentes dos estudos das variações de registros, pode-se desmistificar visões estereotipadas que são difundidas principalmente, no âmbito do ensino de língua inglesa.

⁶³What makes a particular expression or speech act situationally appropriate is not so much the linguistic form or the range of the L2 speaker’s linguistic repertoire, but the sociocultural variables, which are rarely addressed in explicit instruction.

2.2.1 Conceituação

Afim de promover um entendimento mais detalhado e claro acerca da interculturalidade, iniciaremos com a conceituação referente ao termo cultura, que embora possa parecer simples, na verdade é complexo, como afirma Nieto (2010, p.135), é um termo que “pode significar coisas diferentes para pessoas diferentes em contextos diferentes.”⁶⁴ Além disso, trata-se de uma palavra com o campo semântico amplo e de uma terminologia com diferentes perspectivas de estudo no campo das ciências da linguagem.

De acordo com Tomalin e Stempleski (1993, p. 15), “quando nós vivemos em um país particular, automaticamente nós nos tornamos expostos e acostumados a uma variedade de símbolos e imagens embutidas em canções, quadros e costumes.” Os autores argumentam que o conhecimento de tais aspectos culturais podem contribuir para o sucesso da comunicação em língua-alvo, logo, dominar uma língua estrangeira implica conhecer a cultura dos seus falantes nativos, por exemplo, no Brasil, a justiça é representada por uma “balança”, mas ao nos depararmos com uma balança em um estabelecimento comercial, não podemos nos referir a ela como “justiça”. Visto que a relação língua-cultura pode influenciar não apenas o modo de falar, mas também as práticas sociais de um povo, faz-se imperativo compreender o que é cultura.

De acordo com Kramsch (2013), a cultura pode ser entendida em duas perspectivas, a saber, a humanística conhecida como a cultura do “grande C”; e a social, também chamada decultura do pequeno “c”. A primeira, nas palavras da autora, refere-se à cultura como “produto de uma alfabetização canônica impressa adquirida na escola, ela é sinônimo de conhecimento geral da literatura e das artes”⁶⁵ (KRAMSCH, 2013, p. 65). Depreende-se que cultura, nessa concepção está relacionada a interesses da elite, pois privilegia os padrões clássicos, valorizando o que é culto e erudito, em detrimento do que é “popular”, atribuindo a este termo sentidos pejorativos, de baixa qualidade e, até mesmo questionando o seu valor cultural. Canclini (2015, p. 21) realça a oposição entre culto e popular, ao afirmar que “o artesanato ia para as feiras e concursos populares, as obras de arte para os museus e bienais.” Vale ressaltar que este trabalho fundamenta-se em uma visão intercultural, que de acordo com Melo (2016, p. 8) “concebe a existência de várias culturas e uma interação entre elas”,

⁶⁴The term “culture” can be problematic because it can mean different things to different people in different contexts.

⁶⁵Culture is the product of a canonical print literacy acquired in school; it is synonymous with a general knowledge of literature and the arts.

portanto faz-se necessário tratar o estudo da linguagem por meio de lentes que sejam capazes de captar as nuances existentes na relação língua-cultura, revelando fenômenos inerentes ao uso dos PV na língua inglesa.

No que se refere à concepção social de cultura, isto é, a cultura do “c” pequeno, Kramersch (2013, p.66) afirma que se trata de uma cultura do cotidiano, pois “inclui os modos dos falantes nativos de se comportar, comer, falar, habitar, seus costumes, suas crenças e seus valores.”⁶⁶ Pode-se perceber que nessa visão, a cultura se torna mais ampla ao englobar a sociedade em sua totalidade, isto é, não privilegiando e restringindo a cultura como um bem destinado a uma minoria, como destaca Nieto (2010, p. 135) “cultura não é apenas o que um grupo de pessoas da elite podem fazer em seu tempo extra, mas há ainda várias ideias conflitantes do que cultura significa de fato na vida cotidiana.”⁶⁷

Nessa direção, abordamos língua e cultura, através de uma perspectiva ampla e cautelosa, para se evitar descrições equivocadas, pois como afirma Lyons:

Com demasiada frequência a correlação entre língua e cultura é feita num nível muito geral, e com o pressuposto tácito ou explícito de que os que falam a mesma língua têm necessariamente que compartilhar a mesma cultura. Tal pressuposto é manifestamente falso com relação a muitas línguas e muitas culturas (2016, p. 247).

Para Corbett (2010, p.4), “cultura é a expressão dos valores e crenças de uma comunidade.”⁶⁸ Esta, para o autor pode ser definida “em termos de idade, gênero, profissão, etnia, classe social, nacionalidade, ou mesmo afiliação a alguma atividade social como esporte, teatro, consumo de literatura etc”⁶⁹ (CORBETT, 2010, p. 4). Logo, pode-se depreender que há vários aspectos que podem influenciar tanto nas práticas sociais como nas práticas linguísticas dos usuários de determinada língua natural. As palavras podem sofrer alterações semânticas impulsionadas por fatores externos à língua, por exemplo, na língua portuguesa, a palavra “caneta” pode se referir a um objeto utilizado para escrita ou a uma jogada em que o jogador passa a bola entre as pernas do adversário em uma partida de futebol. Na língua inglesa, o mesmo drible é chamado de *nutmeg*, palavra que também é usada para se referir à noz-moscada, tais convenções de sentidos evidenciam a teoria saussuriana da arbitrariedade dos signos linguísticos, além de deixarem clara a influência da cultura no

⁶⁶*It includes the native speakers' ways of behaving, eating, talking, dwelling, their customs, their beliefs and values.*

⁶⁷*Culture is not just what an elite group of people may do in their spare time, but there are still various and conflicting ideas of what culture actually means in everyday life.*

⁶⁸*Culture is the expression of the values and beliefs of a community*

⁶⁹*[...] may be defined in terms of age, gender, profession, ethnicity, social class, nationality, or even affiliation to some social activity such as sport, theatre-going, consumption of literature, and so on.*

processo de construção de sentidos. Desta forma, acreditamos que, ao unir os preceitos da LC aos preceitos da Interculturalidade, podemos elucidar aspectos relevantes inerentes ao comportamento dos PV, em contextos da escrita e da oralidade e em situações formais e informais do uso da linguagem.

Silva (2017, p. 232) descreve a cultura através da dicotomia visível x invisível, segundo o ponto de vista do autor, a cultura visível pode “incluir os estilos de vestir, a culinária, os costumes, os festivais, e outras tradições [...]. O termo cultura invisível se aplica as crenças socio culturais e pressupostos de que a maioria das pessoas nem sequer estão conscientes e, portanto não podem examinar intelectualmente. Nesse sentido, pode-se comparar a cultura a um *iceberg*, pois somente visualizamos uma parte de sua constituição, fato que revela a possibilidade de se encontrar aspectos surpreendentes acerca dos usos dos PV, uma vez que a sua descrição pode revelar aspectos que nem mesmo os falantes nativos estão conscientes.

De acordo com Risager (2006, apud MELO, 2016, p. 4), cultura pode ser definida de modo universal e de modo mais específico, já que há traços comuns a todas as culturas e aspectos particulares, por exemplo, em todas as culturas há a presença de um líder, porém não é qualquer cultura que permite que a mulher exerça esse papel. No que tange aos PV, são comuns na língua inglesa, mas não possuem estruturas gramaticais semelhantes na língua portuguesa, tal fato também revela que, apesar de usarmos línguas naturais para nos comunicar, elas compartilham certos aspectos, mas se diferem em outros, como evidenciado por Chomsky ao propor a existência de princípios universais e parâmetros particulares, vale ressaltar, que apesar de não nos inscrevermos na teoria gerativista cuja base é alicerçada na mera intuição dos linguístas, nós não descartamos radicalmente todos os seus pressupostos, uma vez que os estudos empíricos da linguagem baseados em *corpora* podem facilmente validar ou refutar tais pressupostos através de estudos que refletem o uso real da linguagem.

Nieto traz uma definição de cultura que revela seu caráter dinâmico, ao abordá-la:

Como valores sempre em transformação, tradições, relações sociais e políticas, e visões de mundo criadas, compartilhadas e transformadas por um grupo de pessoas ligadas por uma combinação de fatores que podem incluir uma história comum, localização geográfica, linguagem, classe social, e religião⁷⁰ (2010, p. 136).

Essa definição permite-nos apontar a escolha de um *cópus monitor* como a mais apropriada, já que pretendemos comparar os dados referentes aos usos dos PV, no trabalho realizado em 2010 com os dados descobertos na presente pesquisa, então como a autora deixa

⁷⁰*I have defined culture as “the ever-changing values, traditions, social and political relationships, and worldview created, shared, and transformed by a group of people bound together by a combination of factors that can include a common history, geographic location, language, social class, and religion.*

claro, há traços culturais que são mutáveis, e a língua também pode sofrer transformações ao longo dos tempos como apontam pesquisas realizadas na sociolinguística, por exemplo.

Independentemente do modo como se dicotomizam os elementos culturais, fica claro que todos os conceitos convergem para a descrição de princípios éticos e morais, de produtos como comidas e vestuários, e de práticas individuais ou coletivas, que traduzem os modos de ser e agir das pessoas, em cotextos diversos e situações variadas, localizadas no tempo e no espaço. Uma vez que as noções de tempo e espaço sofreram mudanças devido aos avanços tecnológicos e às demandas da globalização, o modo de se lidar com as linguagens e com as culturas precisou se adaptar, assim, surgiram termos como multicultural – evidenciando a existência de várias culturas; transcultural – realçando os atravessamentos culturais; e intracultural e intercultural – os quais indicam a necessidade de se lidar com os conflitos ocasionados pela permeabilidade cultural presente em nossas comunidades.

Visto que a linguagem é, não apenas veículo, mas também produto da cultura, isto é, a língua dissemina a cultura, ao passo que é influenciada pelos símbolos e significados culturalmente produzidos pelo convívio social, entendemos que é imprescindível para este estudo, fazer ressoar em seus resultados, a relação língua, cultura e sociedade, produzindo dados que reflitam padrões de comportamento reais dos falantes nativos da língua inglesa em sua variedade norte-americana, à medida que geramos conhecimento atualizado ao descrever os usos relacionados aos PV e, acima de tudo, conhecimento que pode contribuir para a superação de problemas gerados na comunicação em língua-alvo, oriundos do desconhecimento da cultura-alvo e do não desenvolvimento da competência comunicativa intercultural, conceito que será explicado, ao passo que trataremos, na seção posterior, da interculturalidade.

Portanto, ao se lidar com a linguagem é essencial considerar seus aspectos culturais, que de acordo com Corbett (2010, p. 4, tradução nossa), “todos os fenômenos culturais geram sentidos que precisam ser interpretados em relação às convenções locais e valores da comunidade específica que os criaram, adotaram e os adaptaram.⁷¹”. Assim, é crucial não tomar tais fenômenos como fixos e estanques, mas como mostra a visão cultural adotada por Nieto, os fenômenos são dinâmicos, isto é, estão em constante movimento, podem se transformar e produzir sentidos diferentes. Nessa arena versátil, não se pode esquecer de que é preciso ter consciência das singularidades do “eu” e dos “outros” a fim de se evitar equívocos causados pela construção de estereótipos e da supervalorização de uma cultura em detrimento

⁷¹*All cultural phenomena generate meanings that have to be interpreted in relation to the local conventions and values of the specific community that has created or adopted and adapted them.*

da outra, pois como afirma Kramersch (2013, p. 60, tradução nossa) “em seu uso, o signo linguístico significa mais do que sua definição dicionarizada.⁷²” e, segundo Corbett (2010, p.3, tradução nossa), no que tange à variação presente nos registros, “muitos gêneros são governados por um conjunto de convenções linguísticas relativamente estáveis que são apropriadas para seu propósito cultural.⁷³” Logo, ao adotar uma lente intercultural para tratar do uso dos PV, abre caminho para descobertas de fenômenos relevantes não apenas a descrição da linguagem, como também, para o enriquecimento de gramáticas, dicionários, e materiais didáticos. O tratamento linguístico proposto por esta pesquisa produz evidências de que os sentidos das palavras não estão isolados em si mesmas, mas são altamente influenciados pelo uso da linguagem e pelas variações de registro.

Melo (2016, p.6) destaca que “a língua é construída pelo uso que cada indivíduo faz dela, e esse uso é carregado de subjetividade e valores e crenças.” Nesse sentido, é importante ressaltar que os indivíduos pertencentes a uma mesma comunidade e falantes de uma língua comum, participam de grupos sociais diferentes e podem compartilhar alguns elementos culturais e não se identificarem com outros, então, percebe-se que os conflitos gerados pelos embates culturais não são restritos aos encontros de falantes de línguas diferentes. Cabe então, definir o termo intercultural, porém, desse cenário dicotômico, emerge também o termo intracultural.

Ao tratar do termo intraculturalidade, Kecskes (2012, p. 68) aponta para as relações entre membros de uma mesma cultura predominante, mas com valores um pouco diferentes ligados às várias subculturas presentes em uma comunidade. Logo, compreende-se que as relações intraculturais revelam a necessidade de se lidar com as diferenças culturais entre falantes de uma mesma língua-cultura. Já, ao definir interculturalidade, o autor a descreve “como um fenômeno que não é apenas interacional e socialmente construído no curso da comunicação, mas também depende de modelos culturais relativamente definíveis e normas que representam as comunidades de fala as quais os interlocutores pertencem⁷⁴” (KECSKES, 2012, p. 69). Logo, percebe-se que a interculturalidade pressupõe que os falantes de língua estrangeira se apropriem não apenas da língua-alvo, mas também conheçam os elementos culturais que permeiam a língua. Melo (2016, p. 7) salienta que “em um encontro intercultural cada um (re) produz sua cultura e, simultaneamente, produz cultura, porque cada sujeito tem

⁷²*In its use, the linguistic sign means more than its dictionary definition.*

⁷³*Many genres are governed by a relatively stable set of linguistic conventions which are appropriate to their cultural purpose.*

⁷⁴*Interculturality is a phenomenon that is not only interactionally and socially constructed in the course of communication but also relies on relatively definable cultural models and norms that represent the speech communities to which the interlocutor belong.*

poder criativo sobre sua cultura, e não a reproduz cegamente.” Assim, é extremamente importante que em uma comunicação intercultural, não haja a supervalorização de uma cultura em detrimento da outra. Nessa direção, Melo (2016, p. 8) define interculturalidade como sendo “uma relação de mão dupla entre culturas que, ao meu ver, precisa de certa harmonia ou capacidade de conviver com diferenças para acontecer.” Portanto, depreende-se que mesmo, em um cenário intracultural, medidas interculturais podem contribuir para a superação de diferenças, evitando a construção de estereótipos. Nesse sentido, Byram (2012,87) ressalta que “uma análise refinada mostraria que diferentes grupos falando aparentemente a mesma língua também têm discursos diferentes, e que a mediação intercultural é também relevante aqui como quando línguas e culturas são visível e obviamente diferentes⁷⁵”, confirmando o pensamento de Melo (2016, p.9) “interculturalidade é uma posição de respeito.”

Segundo Kramsch (2012, p. 212), “o contato intercultural se refere a um estado de relações que ocorrem quando pessoas de culturas diferentes entram em contato⁷⁶”, ela ainda acrescenta que “comunicação intercultural ou transcultural é um campo de pesquisa que estuda como as pessoas entendem umas as outras além dos grupos de fronteiras de vários tipos: nacionais, geográficas, étnicas, profissionais, de classe ou de gênero⁷⁷” (KRAMSCH, 2001, p. 201). Nota-se, desse modo, que a cultura tem um papel preponderante na comunicação, pois a linguagem se apóia na cultura para produzir sentidos, logo, fica evidente que para dominar uma língua e se comunicar de forma eficiente é necessário conhecer não só a língua, isto é, os códigos linguísticos, seus sentidos e suas estruturas sintáticas, como também a cultura que permeia a língua-alvo, nesse sentido, Byram (2012, p. 88) defende a importância de se desenvolver a competência comunicativa intercultural que, segundo o autor, “combina o uso em décadas recentes do conceito da ‘competência comunicativa’ em outra língua – com ênfase na habilidade de usar uma língua não apenas com correta aplicação do conhecimento gramatical, mas também em modos socialmente apropriados⁷⁸”, isto é, associando a competência comunicativa à competência intercultural. Esta perspectiva

⁷⁵*A fine-grained analysis would show that different groups apparently speaking the same language also have different discourses, and that intercultural mediation is just as relevant here as when languages and cultures are visibly and obviously different.*

⁷⁶*Intercultural contact refers to a state of affairs that occurs when people from different cultures come in touch with one another.*

⁷⁷*Intercultural or cross-cultural communication is an interdisciplinary field of research that studies how people understand each other across group boundaries of various sorts: national, geographical, ethnic, occupational, class or gender.*

⁷⁸*Combines the use in recent decades of the concept of ‘communicative competence’ in another language – with emphasis on the ability to use a language not only with correct application of knowledge of its grammar but also in socially appropriate ways – with ‘intercultural competence’.*

converge com os preceitos defendidos por Nation, que ressaltam a necessidade de se conhecer a forma, os sentidos e os usos das palavras ao se lidar com uma língua. Visão sintetizada no quadro a seguir:

Quadro 4 – *What is involved in knowing a word*

Form	Spoken	R	What does the word sound like?
		P	How is the word pronounced?
	Written	R	What does the word look like?
		P	How is the word written and spelled?
	Wordparts	R	What parts are recognizable in this word?
		P	What word parts are needed to express meaning?
Meaning	Form and meaning	R	What meaning does this word signal?
		P	What word form can be used to express this meaning?
	Concept and referents	R	What is included in this concept?
		P	What items can the concept refer to?
	Associations	R	What other words does this make us think of?
		P	What other words could we use instead of this one?
Use	Grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	Collocations	R	What words or types of words occur with this one?
		P	What words or types of words must we use with this one?
	Constraints on use (register, frequency ...)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when, and how often can we use this word?

Note: In column 3, R = receptive knowledge, P = productive knowledge. Fonte: Nation (2001, p. 27)

Ao tratar de aspectos importantes que envolvem o conhecimento da linguagem, Nation (2001, p. 323, tradução nossa) afirma que “a maior parte da linguagem que usamos consiste de combinações familiares. Apenas uma minoria é inteiramente nova.”⁷⁹ Então, é essencial conceber o estudo de palavras que contemple não apenas sua ocorrência, mas acima de tudo sua co-ocorrência, isto é, como elas se agrupam para produzir determinados sentidos em contextos e registros diversos, em outras palavras, é relevante lidar com conceitos de colocação, variação e registro.

No que tange à colocação e à frequência, Nation (2001, p. 325, tradução nossa), afirma

⁷⁹*Most of the language we use consists of familiar combinations. Only a minority is entirely new.*

que “se a frequência da colocação é alta e ela ocorre em muitos usos da linguagem, ela merece atenção.⁸⁰” Assim, depreende-se que conhecer as palavras mais recorrentes, as estruturas onde elas tendem a ocorrer, em quais registros e em quais contextos elas são utilizadas no uso real da linguagem pode produzir aportes consideráveis para a formação de falantes não apenas fluentes, mas também interculturalmente competentes.

Além do conhecimento das formas e de seus sentidos, faz-se imperativo ter conscienciadas forças restritivas que regulam o uso adequado da linguagem. Nesse sentido, segundo Nation (2001, p. 57, tradução nossa), “há vários fatores que limitam onde e quando certas palavras podem ser usadas. Falhas em observar estes podem resultar em uso inapropriado.⁸¹” Ele ainda acrescenta que “muitas restrições acerca do uso são melhor tratadas pela discussão e comparação transcultural explícita.⁸²”(NATION 2001, p. 58, tradução nossa), fato que justifica nossa escolha pela abordagem intercultural associada aos estudos dos registros e da variação que possibilita um olhar diacrônico sobre os fenômenos da linguagem e a comparação dos dados em diferentes registros, produzindo conhecimento relevante para o desenvolvimento científico da área e para a renovação de materiais pedagógicos adequando-os às exigências dos documentos que regulam a educação básica e superior no Brasil.

De acordo com Biber e Conrad (2009, p. 9, tradução nossa), “todos os falantes usam a língua em diferentes contextos, sob circunstâncias diferentes, para propósitos diferentes.⁸³” Esse pressuposto pode ser evidenciado pela análise do quadro, pois Nation destaca pontos fundamentais para tornar os estudos lexicológicos mais precisos, uma vez que os falantes utilizam a língua por meio de escolhas léxico-gramaticais sistematizadas, isto é, padronizadas e impulsionadas pela frequência. As escolhas não ocorrem de modo aleatório evidenciando marcas da subjetividade do falante. É a função social e comunicativa que define as estruturas linguísticas que compõem a organização da mensagem. A colocação revela que determinadas unidades lexicais tendem a se relacionar de modo restrito com outras palavras. Embora outros emparelhamentos sejam possíveis, eles não ocorrem no curso natural da linguagem. Os sentidos expressos pelas palavras extrapolam seus próprios limites, sendo fortemente afetados pelo contexto em diversas situações vivenciadas pelos falantes que lançam mão de diversos registros para satisfazerem suas necessidades comunicativas.

Os registros possuem traços linguísticos específicos, atuam em situações particulares e

⁸⁰*If the frequency of a collocation is high and it occurs in many different uses of the language, it deserves attention.*

⁸¹*There are several factors that limit where and when certain words can be used. Failure to observe these can result in inappropriate use.*

⁸²*Most constraints on use are best dealt with by discussion and explicit cross-cultural comparison.*

⁸³*All speakers use language in different contexts, under different circumstances, for different purposes.*

desempenham funções comunicativas culturalmente estabelecidas. Segundo Biber e Conrad (2009, p. 8, tradução nossa), “as características de qualquer registro individual se tornam mais aparentes quando ele é comparado a outros registros.⁸⁴” Logo, lidar com os PV em uma perspectiva capaz de trazer à tona aspectos culturais de seus usos, revelando comportamentos em registros distintos, proporciona condições favoráveis para descobertas relevantes para o uso consciente destas lexias complexas no âmbito intercultural.

O quadro traz questionamentos (eles serão abordados com maior profundidade na próxima seção) que podem ajudar os usuários de línguas estrangeiras a se tornarem ‘falantes interculturais’ que de acordo com Byram (2012, 89), demonstram “que a competência intercultural é valiosa em si mesma e não deve ser considerada uma pobre imitação da competência de um falante nativo⁸⁵”, pois nem todo falante nativo sabe se comunicar harmoniosamente dentro de sua própria cultura e não é capaz de enxergar a pluralidade cultural que o cerca, conhecimento que é essencial para se transitar e se comunicar confortavelmente no mundo contemporâneo, ao invés de se tornar uma ilha rodeada por outros, mas isolada em si.

Nesse sentido, Byram (2012, 92) advoga pela necessidade de se formar cidadãos interculturais, ao afirmar que “vão além disso, envolvendo ambas, a atividade com outras pessoas no mundo e as competências requeridas para o diálogo com pessoas de outras *langacultures*⁸⁶”. Desse modo, o falante transcende seus limites, ao desenvolver uma cidadania global, sem se desvincular de seus valores locais, tornando-se capaz de interagir verbal e não verbalmente, em espaços micro e macroculturais.

Portanto, o domínio da competência intercultural, não se restringe apenas às necessidades de comunicação entre falantes de línguas e culturas diferentes, mas também, é fundamental para se construir relações sólidas dentro de comunidades que compartilham a mesma língua-cultura. Fato que evidencia a necessidade de inserção de práticas pedagógicas interculturais tanto no ensino de língua materna quanto no ensino de língua estrangeira.

Os PV são unidades lexicais muito idiomáticas, isto é, não basta apenas interpretar os sentidos do verbo lexical e associá-los aos sentidos expressos pelas preposições e advérbios que os compõem. Aspecto que permite que se agrupem formando estruturas linguísticas específicas e variadas, em diversas situações da linguagem e exercendo funções

⁸⁴*The characteristics of any individual register become more apparent when it is compared to other registers.*

⁸⁵*To indicate that intercultural competence is worthwhile in itself and should not be considered a poor imitation of native speaker competence.*

⁸⁶*Intercultural citizenship goes beyond this, involving both activity with other people in the world and the competences required for dialogue with people of other languacultures.*

comunicativas em registros da oralidade e da escrita. Como atuam em diversos contextos culturais, abordá-los em uma perspectiva intercultural permite ao pesquisador descrever inúmeros fenômenos relacionados aos seus comportamentos sincrônicos ou diacrônicos revelando características de usos reais da linguagem e as variações ocorridas dentro dos registros estudados, cujos conhecimentos podem contribuir para a evolução do campo científico e para a área do ensino de línguas. As relações entre interculturalidade e ensino serão tratadas na próxima seção.

2.2.2 *Interculturalidade e Ensino*

Sarmento (2004, p. 10) ao tratar da relevância do papel que os fatores sociais e culturais exercem na comunicação, revela as contradições existentes no ensino de línguas, quando afirma que “muito mais tempo do ensino formal parece ser dedicado apenas ao treinamento de aspectos gramaticais em detrimento dos aspectos sociolinguísticos.” Desse modo, a autora evidencia a necessidade de se preencher lacunas de conhecimento fundamentais não apenas para melhorar a qualidade do ensino de línguas estrangeiras, mas, acima de tudo, para desenvolver a competência comunicativa intercultural dos aprendizes, pois, como afirma Hinkel (2014, p. 395), “sem instrução e entendimento das normas culturais e sociopragmáticas da L2, aprendizes não têm e não podem fazer escolhas essenciais necessárias para otimizar sua competência comunicativa.⁸⁷” Portanto, a fim de que o ensino de línguas se torne realmente significativo para os estudantes, capacitando-os para realizar interações sociais adequadamente, ao passo que encontram soluções para os problemas do cotidiano através do uso da linguagem, é de suma importância que o ensino de línguas contemple as relações existentes entre língua, cultura e sociedade. Estas relações podem ser melhor investigadas sob a luz da análise de registros, que fundamenta-se no tripé: aspectos linguísticos, situacionalidade e funcionalidade. Nesse sentido, Biber e Conrad (2009, p. 7, tradução nossa) afirmam que uma análise de registro busca interpretar “o relacionamento entre as características situacionais e traços linguísticos dominantes em termos funcionais.⁸⁸” Fato que nos permite lidar com a variação de registros relacionadas ao uso dos PV.

Ao mencionar o principal benefício em longo prazo do ensino de línguas associado ao ensino de cultura, Hinkel (2014, p. 397, tradução nossa) afirma que “pode-se prover os

⁸⁷*Without instruction in and an understanding of L2 cultural and sociopragmatic norms, learners do not have and cannot make the essential choices needed to optimize their communicative competence.*

⁸⁸*[It interprets] the relationship between situational characteristics and pervasive linguistic features in functional terms.*

aprendizes com a consciência e as ferramentas que lhes permitirão alcançar seus objetivos acadêmicos, profissionais, sociais e pessoais e torná-los bem sucedidos em seu uso diário em ambientes de língua-alvo.⁸⁹” Assim, percebe-se que se faz necessário conhecer as palavras de forma plena, que de acordo com Nation (2001, p.26) “no nível mais geral, conhecer uma palavra envolve forma, sentido e uso⁹⁰”, e Hinkel (2014, p.398, tradução nossa) acrescenta que “na aprendizagem acerca do impacto nas normas socioculturais sobre o uso da linguagem, o primeiro passo é entender que elas existem em todas as línguas, incluindo a L1 dos aprendizes⁹¹” Nesse sentido, a autora ainda argumenta, em Hinkel (2010, p. 396, tradução nossa) que “a complexidade do ensino de culturas repousa no fato de que muitas pessoas que se engajam em interações trans-culturais não são conscientes do impacto indelével da cultura invisível – da própria e da dos outros – em praticamente todos os usos sociais da língua.⁹²” O pensamento da autora evidencia que ao fazer uso da linguagem, acionamos inconscientemente elementos, valores e crenças que influenciam em nosso modo de falar, nas escolhas léxico-gramaticais e nos sentidos transmitidos, portanto compartilhar esses valores culturais em uma comunicação entre falantes de línguas diferentes, pode contribuir substancialmente para o sucesso da comunicação.

No quadro4, apresentado na seção anterior, Nation, ao tratar do uso da linguagem e suas relações com domínio do vocabulário, destaca duas questões abordando as funções gramaticais, uma ligada à recepção da linguagem “Em que padrões a palavra ocorre?”⁹³” (Nation, 2001, p. 27, tradução nossa) e outra à produção de linguagem “Em que padrões devemos usar esta palavra?” (*ibidem*). Uma possível resposta satisfatória seria que de acordo com a intencionalidade do falante, certas construções seriam selecionadas em detrimento de outras para a modalização dos sentidos, por exemplo, *must*+ verbo+ complemento poderia ser usada para expressar uma idéia de obrigatoriedade e prescrição, enquanto *should* + verbo + complemento poderia ser a estrutura empregada para expressar uma sugestão e outras estruturas poderiam ser articuladas para variar entre o nível de polidez da sentença. Um falante nativo internaliza estas construções naturalmente ao fazer uso da linguagem no seu dia

⁸⁹*The most important long-term benefits of culture teaching may be to provide learners with the awareness and the tools that will allow them to achieve their academic, professional, social and personal goals and become successful in their daily functioning in L2 environments.*

⁹⁰*At the most general level, knowing a word involves form, meaning and use.*

⁹¹*In learning about the impact of sociocultural norms on language use the first step is to understand that they exist in all languages, including learners' L1.*

⁹²*The complexity of teaching cultures lies in the fact that most people who engage in cross-cultural interactions are not aware of the indelible impact of the invisible culture – their own and that of other participants – on practically all social uses of languages.*

⁹³*In what patterns does the word occur? In what patterns must we use this word?*

a dia, isto é, inconscientemente ele é influenciado pela sua língua-cultura.

No mesmo quadro, o autor trata dos colocados, isto é, as palavras que frequentemente se associam a outras palavras, as quais também podem ser impulsionadas pela cultura, principalmente se considerarmos as metáforas, por exemplo, *turn into a pumpkin*, formada pelo emparelhamento de um *phrasal verb+artigo+pumpkin*, nesse caso, não foi usado um vegetal qualquer, pois a abóbora é a única escolha possível para expressar a ideia de ir para a cama, ou voltar para casa, pois retoma um elemento cultural relacionado à literatura, mais precisamente a um conto de fadas mundialmente conhecido, a saber, Cinderela. Nesse sentido, ao tratar das relações semânticas do léxico, Melo (2016, p. 6) destaca que “uma língua carrega consigo elementos culturais, como léxico, sentido das palavras (conotativo e denotativo) e também por isso é preciso pensar a cultura no momento de ensinar a língua.” No caso da Cinderela, trata-se de um elemento cultural mundialmente disseminado, tornando essa expressão, possivelmente mais previsível, porém quanto mais restrito a uma língua-cultura, for o traço cultural, menos previsível será a compreensão da metáfora.

Ainda, com relação ao quadro 4, o autor também menciona as pressões que as palavras recebem dos registros, questionando quando, onde e com qual frequência são empregadas na comunicação.

Nesse sentido, o COCA oferece a possibilidade de se lidar com inúmeros registros distribuídos nas oito seções que o compõem. Os registros, de acordo com Berber Sardinha (2021) podem ser classificados como mais oralizados ou mais letrados, estes se referem aos usos da linguagem em sua variedade padrão, enquanto aqueles se referem aos usos que se distanciam da língua padrão. Percebe-se que dominar o uso da linguagem não requer apenas conhecer quais padrões linguísticos são mais formais ou menos formais, é necessário empregá-los apropriadamente a fim de garantir a eficácia da comunicação e evitar constrangimentos e aborrecimentos ligados ao uso equivocado da língua-cultura.

Byram (2012, p. 88) afirma que “aprendizes também adquirem um modo sistemático de desenvolver competências interculturais através da concentração nos relacionamentos entre sua própria língua-cultura e uma outra⁹⁴”, portanto um ensino intercultural pressupõe a valorização tanto do “nós” como do “eles”, isto é, não há em uma relação intercultural uma cultura superior. Nesse sentido, Sarmiento (2004, p. 5) afirma que “uma cultura deve ser estudada em termos de seus próprios significados e valores, e não de uma maneira etnocêntrica, que é o julgamento de outras culturas a partir da nossa própria cultura” e

⁹⁴*Learners also acquire a systematic way of developing intercultural competences by concentrating on the relationships between their own languaculture and one other.*

Kovalec (2013, p. 784) acrescenta que “o processo de interação com a língua-cultura do outro deve ser feito de maneira cuidadosa, a fim de que não ocorra a supervalorização de uma cultura em detrimento da outra.”

Portanto, uma abordagem intercultural para o ensino de língua estrangeira pode contribuir para desenvolvimento da consciência de sua própria cultura, por meio da comparação de ambas as culturas, promove a empatia e o respeito, ao passo que pode garantir a formação de falantes que dominem não apenas as formas linguísticas e seus sentidos, mas que acima de tudo, saibam quando, onde e por que usá-las, produzindo mensagens claras e adequadas em língua-alvo. Nesse sentido, Byram (2012, p. 87, tradução nossa) afirma que a “competência intercultural é crescentemente usada como um conceito para planejar e avaliar a aprendizagem e o ensino quando a mudança mais geral para usar competências tem ocorrido por meio da educação formal.⁹⁵” Nessa direção, planejar reavaliar o processo de ensino-aprendizagem de língua estrangeira em uma perspectiva intercultural requer objetivos coerentemente alinhados e que desafiem os aprendizes a não simplesmente memorizar listas de palavras e estruturas léxico-gramaticais, é preciso capacitá-los para o uso consciente da linguagem, ao passo que compreendem que as escolhas léxico-gramaticais associadas aos aspectos extra-linguísticos produzem sentidos distintos que tornam a linguagem mais polida ou mais rude, mais formal ou menos formal, por exemplo.

Portanto, diante de todos os fatos apresentados até aqui, fica evidente que a cultura desempenha um papel essencial no emprego da linguagem. Então, para se lidar com a linguagem de maneira satisfatória, é necessário compreender que há forças que atuam tanto no interior quanto no exterior do sistema linguístico e que se não forem devidamente consideradas, as descrições linguísticas não conseguirão alcançar a extensão esperada, pois compreender como as palavras são selecionadas, como elas se relacionam, em que registros são mais presentes é crucial para revelar características ligadas ao seu uso cotidiano, levantando dados linguísticos relevantes tanto para a evolução dos estudos científicos quanto para a conscientização de falantes de línguas estrangeiras. Assim, o presente trabalho pretende apresentar resultados acerca dos usos dos PV, que reflitam o comportamento real dos falantes nativos, ao passo que promovam o estreitamento entre a teoria e a prática no ensino de língua inglesa, através da comparação de dados em diversos registros da linguagem, trazendo à luz, traços inerentes à variação encontrada nos diversos usos da linguagem.

⁹⁵*Intercultural competence is increasingly used as a concept to plan and evaluate learning and teaching as the more general change to using competences has taken place throughout formal education.*

3 MATERIAL E MÉTODO

Apresentamos na sequência as seções de material e método.

3.1 Material

Os materiais utilizados para o desenvolvimento de nosso estudo foram os seguintes: A lista de PV de alta frequência (LIU, 2011), *Corpus of Contemporary American English*, Documentos norteadores da educação superior e do Curso de Letras e os dicionários. Todos são descritos nas próximas subseções.

3.1.1 Lista dos 150 *Phrasal Verbs* de alta frequência (Liu, 2011)

Phrasal verbs apresentam desafios não apenas pela abundância e polissemia, mas também pelos obstáculos encontrados no processo de identificação e principalmente por se apresentarem em quatro formas distintas.

Na pesquisa realizada em 2011, Liu afirma não ter realizado nenhuma busca por PV em que o verbo aparece separado de sua partícula por três ou mais palavras, porque “são raros e a busca por eles resultarão em muitos falsos PV⁹⁶” (GARDNER; DAVIES, 2007, p. 344-345 apud LIU, 2011, p. 665). A lista com os PV e suas respectivas frequências por milhão está disponível no apêndice A.

Na pesquisa, Liu (2011) lançou mão de uma abordagem *cross-register* para descrever os padrões de uso dos PV nos cinco diferentes registros do COCA naquela oportunidade. Além disso, seu percurso metodológico permitiu acrescentar novos PV de alta frequência aos estudos anteriores, a saber, Biber *et. al.* (1999), uma lista de 31 PV; e à lista de Gardner e Davies (2007), uma lista de 100 PV de alta frequência extraída de estudos do BNC (*British National Corpus*), dos 31 extraídos de Biber *et. al.*, 27 já constavam na lista de Gardner e Davies, portanto, tratava-se de fato de 104 PV. A busca adicional realizada por Liu elencou mais 48 PV de alta frequência, resultando então em uma nova lista de 152 PV, porém os PV dos grupos *look around – look round*, *turn around – turn round* foram considerados como apenas 2 PV, chegando, então, à lista de 150 PV de alta frequência nas variedades britânica e norte-americana da língua inglesa, uma vez que as buscas foram realizadas nos dois *corpora*, BNC e COCA.

Os PV que serviram como ponto de partida para nosso estudo são os seguintes:

⁹⁶ *PVs so used are rare, and a search for them will yield “many false PVs.*

go on; pick up; come back; come up; go back; find out; come out; go out; point out; grow up; set up; turn out; get out; come in; take on; give up; make up; end up; get back; look up; figure out; sit down; get up; take out; come on; go down; show up; take off; work out; stand up; come down; go ahead; go up; look back; wake up; carry out; take over; hold up; pull out; turn round/around; take up; look down; put up; bring back; bring up; look out; bring in; open up; check out; move on; put out; look round/around; catch up; go in; break down; get off; keep up; put down; reach out; go off; cut off; turn back; pull up; set out; clean up; shut down; turn over; slow down; wind up; turn up; line up; take back; lay out; go over; hang up; go through; hold on; pay off; hold out; break up; bring out; pull back; hang on; build up; throw out; hang out; put on; get down; come over; move in; start out; call out; sit up; turn down; back up; put back; send out; get in; blow up; carry on; set off; keep on; run out; make out; shut up; turn off; bring about; step back; lay down; bring down; stand out; come along; play out; break out; go round /around; walk on; get through; hold back; write down; move back; fill out; sit back; rule out; move up; pick out; take down; get on; give back; hand over; sum up; move out; come off; pass on; take in; set down; sort out; follow up; come through; settle down; come around/round; fill in; give out; give in; go along; break off; put off; come about; close down; put in; set about.

Trata-se de lista longa para ser estudada e que pode se estender consideravelmente, caso seja levado em conta o campo semântico dos PV, pois podem ser idiomáticos – como se pode depreender das palavras de Gardner e Davies (2007, p. 353, tradução nossa), que lidaram com os sentidos dos PV de alta frequência do inglês britânico, “as mesmas 100 formas que compõem metade de todos os PV no BNC expandiram para 559 sentidos potenciais.⁹⁷” Este fato demonstra que os PV representam um campo rico e desafiador para a realização de pesquisas da LC.

A seguir, apresentaremos o COCA, a fonte de dados do presente estudo.

3.1.2 *Corpus of Contemporary American English*

O COCA foi desenvolvido pelo professor Mark Davies, pesquisador ligado à *Brigham Young University*, no ano de 2008. É um *corpus* que já nasceu gigante, pois contava com mais de 400 milhões de palavras, distribuídas em cinco seções devidamente equilibradas, a saber, *spoken, fiction, popular magazines, newspapers* e *academic journals*. Mas, o fato realmente marcante, não está diretamente relacionado à sua dimensão, e sim, à finalidade de sua concepção. Como já foi explicado no capítulo reservado à fundamentação, há diversos tipos de *corpora*, neste momento, é relevante recapitular os conceitos de *corpora* estáticos e dinâmicos, pois diferentemente dos *corpora* que o precedem, o COCA não tem o objetivo de

⁹⁷ *The same 100 forms that make up roughly half of all PVs in the BNC have expanded to 559 potential meanings.*

apenas representar a língua inglesa em determinada variedade restrita a determinado período de tempo, portanto trata-se de um *cópus* monitor, isto é, tem a missão de proporcionar amostras linguísticas, partindo do momento que compreende o início dos anos 1990 até o presente. Logo, ele vem sendo ampliado anualmente por cerca de 25 milhões de palavras, respeitando os critérios de sua criação, para garantir o equilíbrio das informações. Essa estrutura permite a realização de diversas pesquisas, seja no âmbito lexical, semântico ou discursivo, tornando-se uma fonte confiável para estudos em diversas áreas da linguística, tais como sociolinguística (estudos da variação e mudança linguística), lexicologia, lexicografia, linguística de *cópus* e linguística aplicada de uma forma geral. Vale lembrar, que sua criação preencheu lacunas que limitavam as pesquisas até então, fato que pode ser depreendido da afirmação de Davies (2011, p. 448) “o fato é que até recentemente, verdadeiramente, não há *corpora* monitores confiáveis de inglês. Há alguns *corpora* que se aproximaram de monitores e que têm sido divulgados como monitor *corpora*.[...] entretanto, cada um deles tem sofrido com falhas.” O quadro a seguir detalha melhor a tipologia do COCA.

Quadro 3 – Tipologias presentes no *Corpus of Contemporary American English* (COCA)

Tipologia	Conceito
Monitor	A composição é reciclada para refletir o estado atual de uma língua. Opõe-se a <i>corpora</i> de amostragem.
Dinâmico	O crescimento e diminuição são permitidos, qualifica o <i>corpus</i> monitor.
Equilibrado	Os componentes (gêneros, textos etc.) são distribuídos em quantidades semelhantes (por exemplo, mesmo número de textos por gênero).
Contemporâneo	Representa o período de tempo corrente.
Falado	Composto de porções de fala transcritas.
Escrito	Composto de textos escritos, impressos ou não.
De língua nativa	Os autores são falantes nativos
De referência	Pode ser usado para fins de contraste com o <i>corpus</i> de estudo

Fonte: Berber Sardinha (2004, p. 20 -21).

Compilar um *cópus* englobando todas essas tipologias foi um empreendimento marcante para a LC. As próximas informações acerca da compilação original, das atualizações e inovações realizadas em sua base de dados demonstram o potencial de suas amostras linguísticas e de suas ferramentas para a realização de estudos da linguagem.

A Figura 1 ilustra a página inicial do COCA no ambiente da *web*.



Fonte: Davies (2020).

Atualmente, o COCA conta com mais de 1 bilhão de palavras, distribuídas em 485.282 textos, agora divididos em 8 partes devidamente equilibradas. Este equilíbrio entre os mesmos registros de tempos diferentes permite a realização de estudos que podem ser bem descritos pela metáfora da “comparação de ‘maçãs’ com ‘maçãs’.” Ao analisar a tabela 1, pode-se perceber que, embora exista a preocupação com equilíbrio, as seções não são totalmente compatíveis, portanto, como as amostras não são exatamente do mesmo tamanho, há a necessidade de aplicação de testes estatísticos para atestar se os dados são realmente significativos ou meros frutos do acaso.

Tabela 1 - Quantidade de tokens nos registros do COCA

Registros	<i>Tokens</i>
Blog	125,496,215
Web	129,899,426
Tv/Movies	128,013,334
Spoken	127,396,916
Fiction	119,505,292
Magazine	127,352,014
Newspaper	122,959,393
Academic	120,988,348
Total de <i>tokens</i>	1,001,610,938

Fonte: Davies (2020).

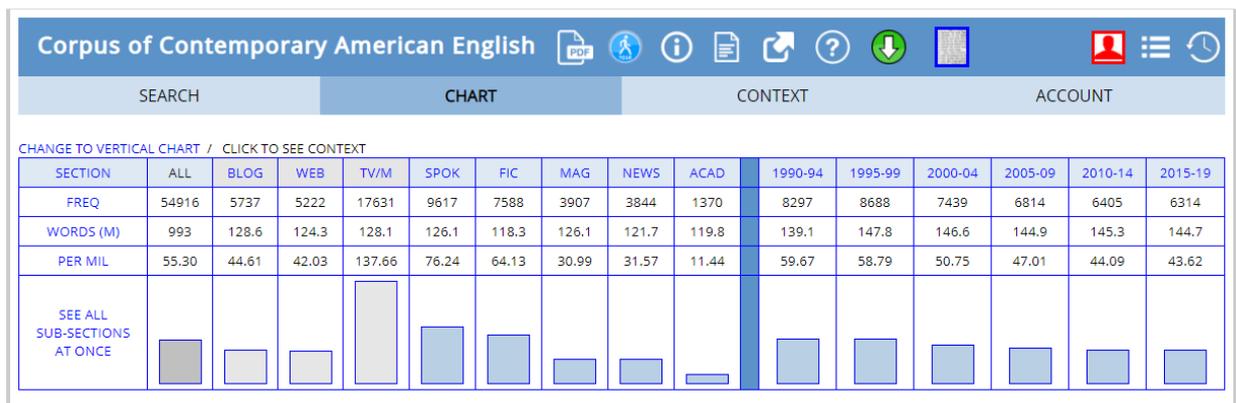
Como afirma Davies (2010, p. 453, tradução nossa), “um aspecto crucial do projeto do COCA é que ele é dividido quase igualmente entre *spoken, fiction, popular magazines,*

newspaper e *academic journals*, 20% em cada gênero⁹⁸”. O córpus e suas ferramentas de pesquisa permitem que o linguísta explore inúmeros aspectos inerentes às mudanças lexicais, morfológicas, sintáticas, semânticas e discursivas.

Ao destacar as mudanças de ordem lexical, Davies (2010, p. 454, tradução nossa) afirma que “em um nível mais básico, o COCA pode dar a frequência de qualquer palavra ou frase em cada um dos períodos de tempo (1990-1994, 1995-1999, 2000-2004, 2005-2009), assim como os cinco principais gêneros⁹⁹”. Após a atualização realizada em março de 2020, vale ressaltar que suas seções foram ampliadas de 5 para 8, tendo sido acrescentadas as seções *blog*, *web* e *tv/movies* e, conseqüentemente, com relação à listagem de períodos apresentada em Davies (2010), foram acrescentados mais dois períodos: (2010-2014 e 2015-2019).

Apresentamos na Figura 02 as frequências de *go on* em um gráfico gerado pelo COCA.

Figura 2 - Frequência do PV *go on* nos registros e nos períodos de tempo



Fonte: Davies (2020).

Para além do estudo de PV, o COCA coloca à disposição do pesquisador, lentes capazes de capturar as mudanças morfológicas, que como afirma Davies (2010, p. 456, tradução nossa) “o COCA pode também ser usado para investigar mudanças na formação de palavras, incluindo a ‘produtividade’ de dados morfemas – onde produtividade se refere a quão livremente um dado morfema pode ser usado para criar novas palavras.¹⁰⁰” A figura 03 mostra a busca por palavras que se associam por meio do hífen à *friendly*, criando novos termos como, por exemplo, *user-friendly*, a mais saliente no córpus, com 1189 tokens.

⁹⁸A crucial aspect of the design of the COCA is that the corpus is divided almost equally between spoken, fiction, popular magazines, newspapers and academic journals, 20 % in each genre.

⁹⁹At the most basic level, COCA can give the frequency of any word or phrase in each of the four times periods (1990-1994, 1995-1999, 2000-2004, 2005-2009), as well as the five main genres.

¹⁰⁰The COCA can also be used to look at changes in word formation, including the ‘productivity’ of given morphemes—where productivity refers to how freely a given morpheme can be used to create new words.

Figura 3 - Busca de **-friendly*

Fonte: Davies (2020).

Outras modalidades de estudos proporcionadas pelo COCA são aquelas direcionadas para a mudança sintática. Nessa direção, Davies (2010, p. 456, tradução nossa) argumenta que “em olhar para a sintaxe, [...] o COCA proporciona muitos dados úteis, que provavelmente, não estariam disponíveis de outras fontes.¹⁰¹” As Figuras 04 e 05 mostram buscas realizadas, a título de exemplificação, as quais apresentam dados relacionados à frequência de *[get]* *[vvn*]*. O comando realiza as buscas e gera listas das coligações entre *get* e verbos no particípio, fornecendo dados referentes às ocorrências nos 8 registros e nos cinco intervalos de tempo.

Esses dados permitem que o linguista perceba tendências de aumento, diminuição ou estabilidade no uso das referidas coligações. É sempre importante ressaltar que a linguagem é de natureza probabilística, isto é, as associações são padronizadas, não derivam do acaso. A LC reforça a credibilidade das pesquisas empíricas em detrimento de estudos pautados em dados especulativos.

Figura 4 - Ocorrência de *[get]* *[vvn*]*

¹⁰¹ In looking at syntax, [...] the COCA provides very useful data, which would likely not be available from any other source.

HELP	CONTEXT	ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 208,091 UNIQUE 8,219 +
1	<input type="checkbox"/>	GET RID	19809	
2	<input type="checkbox"/>	GET MARRIED	8717	
3	<input type="checkbox"/>	GET STARTED	6286	
4	<input type="checkbox"/>	GETTING MARRIED	5569	
5	<input type="checkbox"/>	GET PAID	5546	
6	<input type="checkbox"/>	GET CAUGHT	5408	
7	<input type="checkbox"/>	GOT MARRIED	4515	
8	<input type="checkbox"/>	GETTING RID	3644	
9	<input type="checkbox"/>	GOT CAUGHT	3228	
10	<input type="checkbox"/>	GET INVOLVED	3154	

Fonte: Davies (2020).

No âmbito da investigação das mudanças semânticas, o COCA disponibiliza listas de frequências que fornecem pistas concretas de quando uma forma linguística tem acepções enfraquecidas, ou até mesmo, quando novos sentidos são a ela associados. Logo, para esclarecer esses fatos, o linguista pode facilmente lançar mão de análises de colocados, como afirma Berber Sardinha (2009, p. 42) “podemos depreender o significado das palavras por meio da forma como elas são usadas. Mesmo palavras polissêmicas podem ter seus sentidos desambiguizados por meio da observação de padrões.” Segundo Davies (2010, p. 460), “um dos aspectos únicos do COCA é que ele tem um *thesaurus* com entradas para mais de 60.000 palavras. Usando esta informação semântica, usuários podem facilmente comparar a frequência das palavras em campos semânticos inteiros através do tempo¹⁰²”.

Ao tratar das inovações realizadas na base do COCA e relacionadas à sua atualização mais recente, que além de ampliar sua amostra linguística, também implantou novas ferramentas de busca, possibilitando a realização de pesquisas mais refinadas. Davies (2020) destaca que o usuário:

[...] pode agora analisar textos inteiros (ex. redações de estudantes ou um artigo de um jornal online), e então ver informações detalhadas sobre as palavras e frases no texto, baseado nos dados do COCA. [...] pode clicar sobre qualquer palavra no texto e ver ampla variedade de informações (definição, tradução, etimologia, pronúncia, imagens, vídeos, palavras relacionadas, colocados, tópicos, *clusters*, linhas de concordâncias, e mais)[...] pode também clicar sobre frases de seu texto e então encontrar frases relacionadas no COCA que o ajudarão a descobrir “exatamente a frase correta” para um dado registro¹⁰³(DAVIES, 2020, tradução nossa).

¹⁰² One of the unique features of COCA is that it has an integrated thesaurus with entries for more than 60,000 words. Using this semantic information, users can easily compare the frequency of words in entire semantic fields across time.

¹⁰³ You can now analyze entire texts (e.g. student compositions or an article from an online newspaper), and then see detailed information on the words and phrases in the text, based on data from COCA. You can click on any word in the text and see a wide range of information (definition, translation, etymology, pronunciation, images, videos, related words, collocates, topics, clusters, concordance lines, and more). You can also click on phrases

Portanto, a ampliação do leque de ferramentas à disposição do pesquisador abre caminhos para uma série de novas pesquisas do âmbito da linguística de *corp*us e possibilitando a descoberta e descrição de fatos linguísticos enriquecedores.

3.1.3 Documentos educacionais norteadores

Uma vez que a presente pesquisa direciona seu fazer científico para produzir dados significativos para o ensino de língua inglesa, tendo como público alvo o professor-formador, é importante conhecer as diretrizes nacionais que norteiam os objetivos e as práticas do Ensino Superior no Brasil. Como o objeto da pesquisa está relacionado ao Curso de Letras, também, faz-se necessário analisar as diretrizes específicas desse curso.

Assim, dois documentos foram analisados nessa pesquisa, a saber, as Diretrizes Curriculares Nacionais para o Curso de Letras (BRASIL, 2002) e as Diretrizes Curriculares Nacionais para a Formação Inicial de Professores para a Educação Básica (BRASIL, 2020), que institui a Base Nacional Comum para Formação Inicial de Professores da Educação Básica (BNC-Formação);

Neste documento (BRASIL, 2020, p. 87), as competências específicas para a ação docente são destacadas e divididas em três dimensões fundamentais, “I) o conhecimento profissional; II) a prática profissional; e III) o engajamento profissional.” Por uma questão de delimitação de objetivos, apresentaremos aqui, aquelas que se relacionam com o presente trabalho.

Na dimensão do conhecimento, destacam-se o domínio dos saberes e como ensiná-los; considerando a dimensão da prática, as ações pedagógicas que conduzem à apropriação do conhecimento e ao desenvolvimento de competências e habilidades estão em relevo; por último, a dimensão do engajamento profissional pressupõe o comprometimento com o próprio crescimento profissional. Assim, a realização de pesquisas, como o presente trabalho, produz pontos positivos marcantes para as três dimensões, pois ao trazer à tona os fenômenos inerentes ao uso dos PV em contextos linguísticos e culturais diversos, não apenas propiciamos condições para ampliar esse conhecimento, como ainda indicamos a que situações do cotidiano e a que tipos de textos tal saber está intrinsecamente relacionado, além de estimular a busca por práticas pedagógicas cada vez mais completas e eficientes. Ademais, encorajamos o compromisso com o sucesso profissional, por meio da conscientização de que

in your text and then find related phrases in COCA, which will help you to find "just the right phrase" for a given genre.

o docente não é apenas um transmissor de conhecimentos que partem de outras fontes, mas acima de tudo, é responsável pela construção de novos saberes.

Considerando os fundamentos da política de formação docente, o documento ressalta a importância da produção de novos conhecimentos oriundos de áreas distintas e sua aplicação no contexto escolar/acadêmico para a transposição de obstáculos que impedem o sucesso das práticas pedagógicas, ao afirmar que:

A inclusão, na formação docente dos conhecimentos produzidos pelas ciências para a Educação, contribui para a compreensão dos processos de ensino-aprendizagem, devendo-se adotar as estratégias e os recursos pedagógicos, neles alicerçados, que favoreçam o desenvolvimento dos saberes e eliminem barreiras de acesso ao conhecimento. (BRASIL, 2020, p. 87).

Nesse sentido, nossa pesquisa busca aproximar a teoria e prática, ao levantar dados que não só, podem contribuir para o sucesso do processo de assimilação de conhecimento, como também produz instrumentos para que o docente seja capaz de divulgar de forma mais natural e clara, em língua inglesa, as descobertas provenientes de seu importante papel de professor-pesquisador.

No Ensino Superior, espera-se que o professor-formador adequa sua prática, suas políticas educacionais às propostas da BNCC, como afirma Brasil (2020, p. 87), um dos princípios relevantes é “a articulação entre teoria e prática para a formação docente, fundada nos conhecimentos científicos e didáticos, contemplando a indissociabilidade entre o ensino, a pesquisa e a extensão, visando à garantia do desenvolvimento dos estudantes”.

Concordamos com a ideia de que deve haver “[...] conexão entre ensino e a pesquisa, [...] uma vez que ensinar requer, tanto dispor de conhecimentos e mobilizá-los para a ação como compreender o processo de construção do conhecimento” (BRASIL, 2020, p. 87).

Finalmente, ao propor o ensino de PV, levando em consideração aspectos que vão além da simples memorização de formas e significados, destacando traços relacionados aos seus usos em contextos diversos, se adequa ao teor do documento que diz que, na dimensão da prática profissional, é competência do docente “propor situações de aprendizagem desafiadoras e coerentes, de modo que se crie um ambiente de aprendizagem produtivo e confortável para os estudantes” (BRASIL, 2020, p. 90).

Quanto às diretrizes específicas do Curso de Letras (BRASIL, 2001, p. 29), o currículo do curso deve ser compreendido como “construção cultural que propicie a aquisição do saber de forma articulada. Por sua natureza teórico-prática, [...] o currículo deve ser constituído tanto pelo conjunto de conhecimentos, competências e habilidades [...]”Ao definir os objetivos, o referido documento acentua a importância da competência intercultural, do

domínio das linguagens em contextos diversos de oralidade e escrita, e do desenvolvimento de relações interpessoais.

Salientamos que nossa investigação vai ao encontro das diretrizes particulares da área, pois procura produzir informações linguísticas acerca do uso dos PV em contextos formais e informais, tanto na linguagem oral quanto na escrita. Portanto, trata-se de conhecimento, isto é, subsídio para dominar não apenas traços léxico-gramaticais da língua inglesa, mas também compreender usos e sentidos que extrapolam os limites linguísticos, conscientizando-os de fatos extra-linguísticos.

As diretrizes ainda apontam para uma grade curricular que contemple conteúdos ligados à linguística e à literatura para desenvolver habilidades específicas. E, ainda, de acordo com as diretrizes:

Devem articular a reflexão teórico-crítica com os domínios da prática – essenciais aos profissionais de Letras, de modo a dar prioridade à abordagem intercultural, que concebe a diferença como valor antropológico e como forma de desenvolver o espírito crítico frente à realidade. (BRASIL, 2001, p. 31).

As confluências existentes entre os dois documentos descritos aqui permitem destacar pontos fulcrais no âmbito de ensino de língua inglesa. Em primeiro lugar, merece destaque a necessidade de tornar teoria e prática cada vez mais interligadas. Em segundo lugar, fazer da universidade um espaço não só destinado à difusão de conhecimento, mas sim lugar de formação das competências e habilidades profissionais, interpessoais e interculturais. Por fim, a responsabilidade pelo autodesenvolvimento. Logo, os dois documentos demonstram que, no âmbito acadêmico, ligar teoria e prática, por meio de um currículo que contemple ao mesmo tempo o desenvolvimento profissional, interpessoal, intrapessoal e intercultural, permeado por atividades curriculares planejadas para garantir aprendizagem significativa e autonomia. Esta, no território da linguística, pressupõe conhecimentos acerca da língua materna e/ou estrangeiras e a consciência de seus usos na oralidade e na escrita em contextos formais e informais.

3.1.4 Os dicionários

3.1.4.1 O Longman Dictionary of Contemporary English

Publicado em 2003, o LODCE é resultado de pesquisas realizadas no *Corpus Network*, uma grande base de dados crescente, que na época contava com 300,000,000 de palavras, cobrindo livros, jornais, revistas e inglês falado. Além de vasculharem a internet em busca de

novos verbetes. Os exemplos que aparecem no dicionário são versões editadas de sentenças dos corpuses. Isto significa que os exemplos do LODCE são mais realísticos que em outros dicionários, portanto os estudantes têm contato com as palavras do jeito que elas são realmente usadas, assim, é perceptível que o dicionário descreve a língua como as pessoas de fato a usam.

3.1.4.2 O Collins COBUILD Advanced Learner's Dictionary

A quinta edição publicada em 2006 do dicionário está também disponível em CD-ROM. A fonte de dados utilizada no desenvolvimento do COBUILD é o *Bank of English*, parte do *Collins Word Web*, na época, a maior coleção de dados deste tipo em qualquer língua, contendo 645,000,000 de palavras. Sua amostra é composta por textos de inglês escrito e falado armazenados em computador para estudos da língua em uso. Aproximadamente 40% de suas fontes vêm do inglês britânico, enquanto 30% dos dados vêm do inglês americano e outros 30% da Austrália, Nova Zelândia e Canadá.

Os textos escritos vêm de jornais, revistas, livros de ficção e de não-ficção, websites, folhetos, panfletos, relatórios e cartas. A fala informal é representada por gravações de conversas casuais do cotidiano, reuniões, entrevistas e discursos, além de transcrições de programas de rádio e TV. Os aspectos de sua compilação permitem que os lexicógrafos lidem com diversos sentidos, seus padrões gramaticais e o modo que as palavras se associam.

3.2 Método

Os procedimentos metodológicos compreenderam nove etapas, descritas a seguir:

3.2.1 *Leitura dos documentos*

A leitura das diretrizes referentes tanto ao Ensino Superior de forma geral quanto ao Curso de Letras especificamente, consiste uma etapa importante por contribuir para o planejamento consciente dessa pesquisa que almejou contribuir diretamente na produção de conhecimento novo para sua área e, indiretamente, buscou atuar na sua prática profissional. Assim, essa etapa foi marcada pela leitura dos documentos e a identificação dos princípios relacionados com o objeto e objetivos da pesquisa.

3.2.2 Consulta à lista “Frequência e distribuição dos phrasal verbs mais frequentes no COCA”

O segundo passo consistiu na análise da lista dos 150 PV e dos resultados encontrados por Liu durante suas investigações realizadas durante seu estudo publicado em 2011.

3.2.3 Familiarização com o COCA

O COCA oferece uma estrutura muito bem concebida, com orientações claras e curtas para facilitar a experiência do usuário/pesquisador. Apesar das ferramentas poderem passar inicialmente uma impressão de complexidade, seu ambiente virtual é bem simplificado. Requer certo nível de conhecimento da língua inglesa, uma vez que esta é a língua não só presente em suas amostras, como também nas janelas apresentadas ao usuário.

O processo de familiarização seguiu as etapas estabelecidas pelo próprio desenvolvedor, que afirma que há quatro formas de realizar as buscas no *corpus*. Primeiramente, foi realizada a busca pelas 1.000 palavras mais frequentes, por meio do menu *BROWSE/RANDOM*. O *corpus* exibe a lista com as palavras ordenadas pela frequência, apresenta opções para pronúncia, vídeo, imagem e tradução além de mostrar sua classe gramatical.

Em seguida, foi realizada a busca individual do verbo [*go*], utilizando a opção *SEARCH*, para a qual o *corpus* forneceu uma lista de frequência lematizada.

Na terceira busca, pela função *ANALISE TEXT*, foi processado um *abstract* de um artigo científico. O *corpus* mostrou uma janela com inúmeras informações sobre o texto em questão, tais como número de *tokens*, as palavras aparecem em três cores diferentes, classificando-as de acordo com a frequência em *LOW*, *MID*, *HIGH*. Ao clicar em qualquer palavra, o *corpus* disponibiliza mais ferramentas, que na ocasião não foram testadas.

Finalmente, a busca [**break**], apresentou uma lista com várias ocorrências, tais como *breaking*, *outbreak*, *daybreak* etc. Ao clicar sobre qualquer palavra da lista, uma nova janela aparece com as concordâncias.

3.2.4 Busca dos PV no COCA (ocorrência por milhão)

Nessa fase, foi necessário acessar a base de dados do COCA para realizar as buscas individualizadas por PV, das frequências por milhão, relativas a cada um dos oito registros que compõem sua amostra linguística. Os dados são fornecidos em forma de gráficos e foram copiados e transferidos para uma planilha do Excel.

3.2.5 *Observação da distribuição nos registros*

Os gráficos gerados pelo COCA, nesta etapa, tiveram papel imprescindível, pois apresentam informações acerca dos PV em cada registro, permitindo verificar quais sobressaem em contextos de oralidade e quais têm destaque nos contextos de escrita, produzindo pistas para perceber usos mais formais ou mais informais.

3.2.6 *Tabulação dos dados*

A sexta fase do estudo requereu a tabulação dos dados encontrados, com a finalidade de facilitar a análise nas etapas seguintes.

3.2.7 *Comparação de uso de PV em 2010 e 2020*

Na sétima etapa, os gráficos referentes ao trabalho de Liu (2011) foram confrontados com aqueles gerados pelas consultas realizadas no COCA, em 2020, com a finalidade de verificar possíveis alterações de comportamento, em cada um dos oito registros que compõem a estrutura de amostragem do córpus.

3.2.8 *Comparação entre os PV mais comuns: USA x Britain*

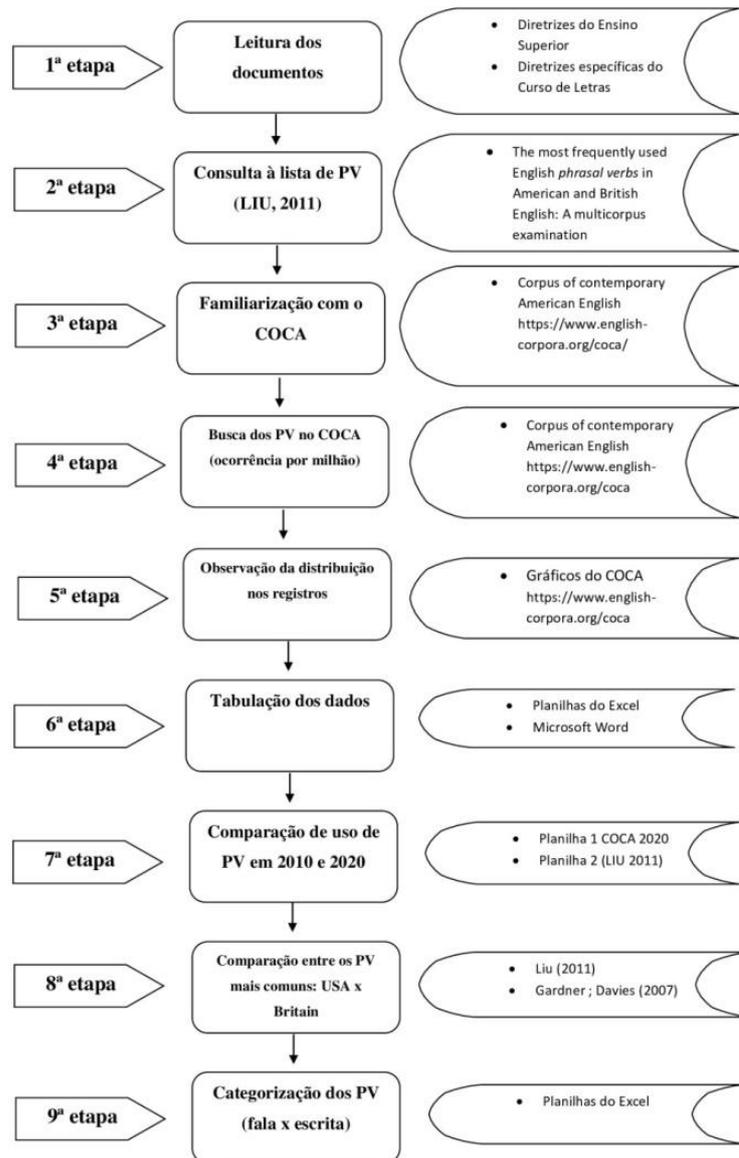
Na oitava etapa, os dados obtidos na pesquisa foram confrontados com a lista fornecida por Gardner e Davies (2007) e os resultados foram tabulados para demonstração.

3.2.9 *Categorização dos PV (fala x escrita)*

Na nona etapa, foi realizada a investigação dos dados que demonstrou quais PV se destacaram da oralidade e quais se destacaram da escrita.

Uma vez que a replicabilidade é um traço importante da tradição científica, e com o intuito de facilitar a compreensão do caminho percorrido nesta pesquisa, adicionamos um fluxograma apresentando suas etapas, descrições e materiais utilizados.

Figura 5 - Fluxograma dos procedimentos metodológicos



Fonte: Próprio autor.

4 RESULTADOS E ANÁLISE DOS DADOS

Nesta seção, descrevemos traços que se destacaram, relacionados aos PV e seus comportamentos em vários registros, considerando não apenas a frequência, mas também, sua distribuição. Primeiramente tratamos dos aspectos inerentes à nova lista, isto é, aos dados coletados das amostras do COCA, em 2020.

Na sequência, realizamos a comparação dos dados encontrados em 2011 com os dados atuais. Na etapa seguinte, analisamos aspectos relacionados às distribuições dos PV nos registros. Depois, buscamos evidenciar tendências de seus usos em contextos tidos como formais e informais. Também nos dedicamos à elucidação dos PV em registros da oralidade e da escrita.

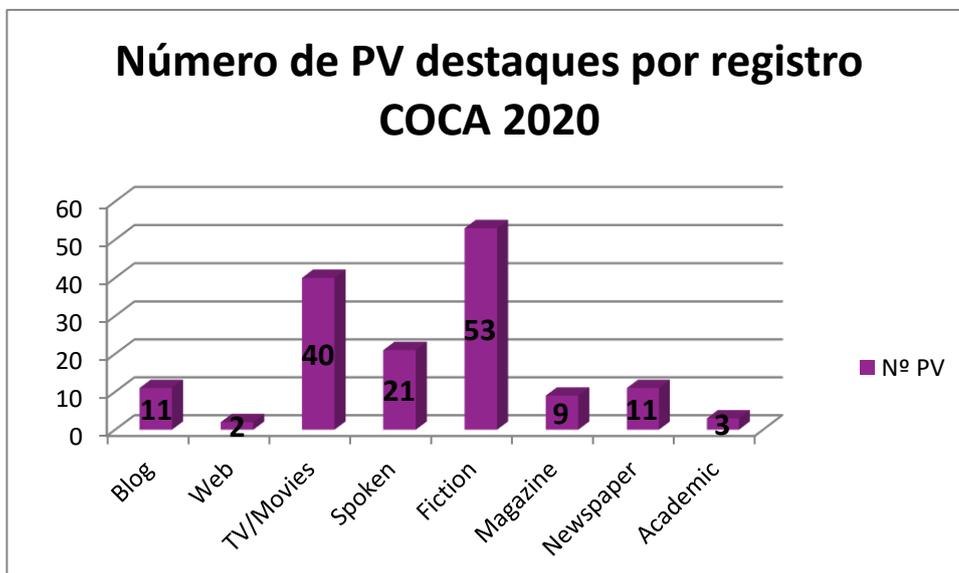
Dedicamos esforços para realizar a comparação entre os PV mais comuns do inglês norte-americano e do britânico, descrevendo sua apresentação em dicionários monolíngues.

De posse de todas essas informações, destacaremos as relações interculturais inerentes aos PV. Finalmente, ao tratar de questões do ensino, consideraremos os documentos oficiais que norteiam as práticas do Ensino Superior, especialmente, aquelas específicas do Curso de Letras, relacionadas ao ensino de língua inglesa, aqui, ligadas aos PV.

4.1 PV no COCA

Para iniciar o detalhamento dos dados extraídos nesta etapa, apresentamos a distribuição nos diferentes registros (orais e escritos) dos 150 PV mais comuns. Os dados estão sintetizados a seguir.

Figura 6 - Presença de PV por registro no COCA, em 2020



Fonte: Dados da pesquisa.

Nota 1: Os números indicam a quantidade de PV diferentes em cada registro.

PV são frequentemente associados ao contexto informal de uso da língua inglesa, pelos leigos, mais um exemplo de especulação e pressuposição intuitiva. Porém, de acordo com informações extraídas do dicionário (LONGMAN, 2003, p. 974), eles são empregados tanto na fala como na escrita, às vezes, seus sentidos podem ser expressos por um único verbo lexical, mas seu uso soa mais formal.

A análise dos dados nos mostra que é preciso ter cautela ao tomá-los como unidades lexicais exclusivas da oralidade e/ou da informalidade, pois eles estão presentes em textos de cada uma das 8 seções do COCA. Porém, alguns são mais salientes nos registros que se referem à fala, seja nos diálogos de TV ou filmes (*TV/movies*), seja nos diálogos de conversas do cotidiano (*spoken*). As duas seções juntas abarcaram 61 dos 150 PV frequentes. Os restantes estão distribuídos em textos dos registros ligados à escrita.

Mais da metade deles (53) destacou-se na seção *fiction*, composta por textos da literatura; onze destacaram-se nos textos de *blogs*, dois nas demais amostras de textos colhidos na *internet*, aninhados na seção *web*, então pode-se considerar que 13 despontam no ambiente virtual. Nove ganham ênfase em revistas (*magazines*), onze em jornais (*newspaper*) e três nos textos de registros acadêmicos. Há PV que servem a propósitos comunicativos relacionados a contextos mais formais, mesmo que em menor número. Moon (2011, p. 53-54, tradução nossa), ao tratar das relações entre multipalavras e registros, afirma que “pelo olhar nas densidades de diferentes tipos de itens de multipalavras em tipos de textos particulares, pode ser visto que há fortes e frequentes preferência por gêneros¹⁰⁴”, estas preferências também ficam evidentes, já nos primeiros passos deste estudo.

O destaque de PV em determinada seção não significa dizer que apresenta frequência por milhão mais alta, que variou, na amostra, entre 1409,49 (*come on*) – na seção *TV/movies* – e frequências que correspondem a menos de 1 ocorrência por milhão, como é o caso de *set about*, também na seção *TV/movies*, com uma (fpm) equivalente a 0.34.

Ao debruçar o olhar sobre os PV, foi possível notar que se pode dividi-los em vários grupos, considerando em qual ou em quais registros eles se destacam. Por exemplo, há aqueles que são mais salientes na fala. Outros, apenas em textos escritos. Alguns, com a

¹⁰⁴By looking at the densities of different kinds of multi-word item in particular text types, it can be seen that there are often strong genre preferences.

chegada da *internet* e toda a gama de registros que compõem a comunicação na *web*, passaram a desempenhar papel relevante na linguagem do mundo virtual.

Certos PV são polivalentes, isto é, atuam na comunicação, em vários registros distintos e desempenham várias funções comunicativas que satisfazem inúmeras necessidades do ser humano, por exemplo, *go on*, que atua em todas as seções do *corpus*, em contextos formais e informais, orais e escritos, este aspecto pode estar ligado ao fato de que *go on* possui 15 acepções diferentes, segundo o dicionário LDOCE. Algumas delas podem aparecer em todos os cenários e contextos de comunicação, por exemplo, “continuar” e “acontecer”; porém há sentidos que atendem a funções mais específicas, como é o caso de “usar como prova” que provavelmente é mais freqüente em contextos de investigação policial.

A seguir, as explicações acerca dos PV seguem os seguintes critérios, 1) *come on* e sua super-ocorrência nos textos televisivos; 2) o comportamento dos dez PV mais frequentes; 3) PV que se destacaram na escrita geral; 4) PV que se destacaram na escrita acadêmica; 5) PV que se destacaram no contexto da *internet*.

Para exemplificar algumas das situações mencionadas, iniciaremos com *come on*, o único que supera a marca das 1.000 ocorrências por milhão. A fpm de 1,409.49 atesta seu uso significativo nos registros da televisão. A adição deste registro na base de dados do COCA, ocasionou mudanças significativas, em relação aos estudos anteriores, pois *come on*, o 25º da lista de LIU (2011) e 14º da lista de Gardner e Davies (2007), simplesmente, saltou para a primeira posição da lista do presente estudo. Esse uso é impulsionado pelos registros da TV e dos filmes, pois na esfera televisiva, ele ocorre 1.409,49 vezes por milhão. No que se refere à fala do cotidiano, ocorre 124,46 por milhão, número muito próximo ao seu uso da ficção, 110,60 por milhão. No contexto acadêmico, sua ocorrência é de apenas 4,83, demonstrando que seu uso é marcante em contextos mais comuns da conversação do cotidiano. Quanto à informação de sua evolução no decorrer das duas décadas, percebe-se que seu emprego continua constante apesar de partir de uma frequência geral de uso no primeiro período (1990-1994), equivalente a 262,79 e no último período (2015-2019), ter apresentado uma queda, chegando a 244,53; e em 2020, de acordo com dados do presente trabalho, 224,94. Demais características de *come on* podem ser observadas no Apêndice B.

No que se refere aos dez mais frequentes, sete se destacam nas seções que refletem registros da conversação, a saber, *go on*, *come back*, *come up*, *go back*, *find out*, *come out* e *go out*. O uso é acentuado nas seções *TV/movies* e *spoken*. De acordo com Liu (2011, p. 662, tradução nossa), ao tratar da relevância dos estudos acerca do comportamento de PV em diferentes registros, “tal informação através de registros, é, entretanto, muito importante para

os propósitos de aprendizagem da língua, porque ela indica os contextos onde *phrasal verbs* específicos são ou não são típicos¹⁰⁵”. Na mesma direção, Gardner e Davies (2007, p. 343, tradução nossa), ao ressaltar a questão do ensino de PV e dos estudos de *corpora*, afirmam que “há muito pouca informação sobre dados reais da frequência que pudesse ser usada por professores, desenvolvedores de materiais e planejadores de testes que desejassem priorizar o tratamento dos PV baseados em seus relativos impactos na linguagem como um todo¹⁰⁶”, portanto, percebe-se que as informações levantadas no presente trabalho podem preencher tais lacunas presentes na área do ensino da língua inglesa.

Go on é mais recorrente na oralidade, embora sua ocorrência seja considerável nos textos da escrita, inclusive em textos acadêmicos, onde apresenta a frequência mais baixa. Fatos que podem contribuir para que seu uso seja tão disseminado. É importante ressaltar que *go on* figura como o PV mais frequente em ambas as listas, a do inglês britânico e do norte-americano. Seu uso vem demonstrando estabilidade nos últimos 20 anos. Seu campo semântico é amplo e demanda esforço para aprendê-lo, portanto sua carga de aprendizagem é pesada, pois, além de polissêmico, está presente em diversos contextos distintos da comunicação. Suas ocorrências mais salientes estão nas seções *TV/movies* e *spoken*, cujas respectivas frequências são 459,04 e 395,35. No que se refere à escrita, destaca-se na seção *fiction*, com ocorrência igual a 228,00, uma frequência considerada alta se compararmos com a maioria das marcas alcançadas por vários PV no COCA. *Go on* ainda se destaca no ambiente da *internet* com frequências de 186,89 na seção *blog* e 171,04 na seção *web*. Em jornais e revistas, seu uso é muito próximo, pois sua frequência em *newspaper* corresponde a 141,22 e em *magazine* equivale a 134,15. Sua ocorrência mais baixa representa seu uso no âmbito da academia com uma frequência total de 6.641 *tokens*, que equivale a sua frequência normatizada igual a 55,44.

Go back e *come back* se destacaram por apresentarem números parecidos em suas respectivas frequências, tanto para a seção *TV/movies* quanto para a *spoken*. Com relação ao primeiro, o gráfico demonstra equilíbrio entre as duas seções da oralidade, pois na televisão sua frequência por milhão é de 196,23 e na fala do cotidiano é de 189,50. Fenômeno semelhante ocorre com o segundo, mas com frequências um pouco mais elevadas, pois no que se refere à televisão, marca 236,74 e na fala do dia a dia, marca 232,61. Ainda, os PV

¹⁰⁵ Such cross-register information is, however, very important for language learning purposes, because it indicates the contexts where specific PVs are and are not typical.

¹⁰⁶ There is very little information about actual frequency data that could be used by teachers, materials developers, and test designers who may wish to prioritize treatment fo phrasal verbs based on their relative impact in the language as a whole.

apresentam em comum, um uso destacado na *fiction*, 228,00 para *come back*; e 150,71 para *go back*. Na *internet*, estão presentes com números, também expressivos, embora bem mais baixos em relação aos da oralidade (consultar apêndice B para visualizar as informações no gráfico do COCA). Por fim, mais um ponto em comum entre os dois PV, é o fato de suas frequências mais baixas informarem sua presença em textos acadêmicos. Na seção *academic*, *go back* marca 18,62; enquanto *come back* traz uma frequência de 11,85.

Come up e *come out* destacaram-se na seção *spoken*, logo as respectivas marcas de 280,52 e 244,70 indicam que são muito recorrentes na oralidade e se consideramos as marcas menos elevadas na seção *TV/movies*, 127,72 para *come up* e 135,73 para *come out*, é evidente que são muito recorrentes na oralidade, segundo as amostras do COCA. Considerando o papel que desempenham na escrita, ambos apresentam números acima de 100 ocorrências por milhão, na seção *fiction*. Ambos têm frequências similares nas seções *magazine* e *newspaper*, todas superiores a 60, apresentam padrões de uso consideráveis na *internet*, sendo que os números de *come out* são um pouco superiores. No que se refere à escrita acadêmica, *come out* atinge a frequência de 19,63; enquanto *come up* alcança 18,91; os números mostram que apesar de desempenharem algum papel nos textos da academia, são muito mais salientes na oralidade. *Come out*, ao se considerar a tendência de uso ao longo das duas décadas que compreendem as amostras do COCA, nota-se que seu uso está aumentando no período de 1990-2019. Já *Come up* apresenta um pico em 2005-2009 (109,03 ocorrências por milhão), mas desde então sofrendo queda na tendência e já se aproxima da sua marca inicial em 1990 (88,92 ocorrências por milhão.)

Ainda tratando dos PV que se destacam na oralidade, *find out* e *go out* se mostraram mais salientes na seção *TV/movies*, o primeiro figura 194,42; enquanto o segundo apresenta frequência equivalente a 153,46. Eles ainda têm papel relevante na seção *spoken*, pois apresentaram números bem chamativos, embora não seja o registro em que obtiveram o maior número de ocorrências. Em terceiro lugar, na seção *fiction*, eles mostram desempenho que aponta um uso bem acentuado, uma vez que *go out* apresenta frequência igual à 112,55 por milhão, e *find out* figura com 87,83 ocorrências por milhão. No que se refere ao papel dos dois PV nos registros da *internet*, *find out* apresenta ocorrência próxima de 100 por milhão, enquanto que *go out* apresenta números um pouco mais modestos nas duas seções, 66,02 (*blog*) e 63,64 (*web*).E, por último, considerando a participação dos dois, no contexto acadêmico, percebemos que suas marcas mais baixas estão na seção *academic*, entretanto, *find out* leva vantagem, provavelmente devido ao fato de sua única acepção ser “descobrir”, adequando-se então ao contexto de inúmeras pesquisas que são realizadas no âmbito da

academia, na busca por inúmeras descobertas. *Go out*, apesar das 12 acepções diferentes dicionarizadas no *Longman dictionary of contemporary English* (LDOCE), apresentou maior marca que *find out*, apenas na seção *fiction*, provavelmente impulsionado pelas acepções que correspondem aos atos de sair de casa, ter um relacionamento romântico e apagar fogo ou luzes.

Grow up traz números surpreendentes, com relação às análises realizadas até agora, pois apresenta uma marca expressiva na seção *spoken* (94,05 *tokens* por milhão) e muito próxima da marca que representa sua participação na seção *newspaper* (96,32 por milhão) apresentando números mais equilibrados entre a oralidade e a escrita. Os números também são muito parecidos, quando tomamos como base de comparação, as marcas encontradas nas seções *TV/movies* (69,05), *web* (67,11) e *blog* (73,24), logo, trata-se de um PV muito versátil, pois ao verificarmos aspectos referentes ao seu uso, é possível depreender que se adéqua a inúmeras necessidades da comunicação humana. Inclusive, pode aparecer em textos acadêmicos, pois possui uma marca, que embora bem mais baixa que as demais, (21,24 por milhão) demonstra que está também presente em textos da academia. Fato que também precisa ser considerado é que vem apresentando um crescimento em sua tendência de uso no decorrer das duas últimas décadas.

Portanto, a análise dos dez PV mais frequentes mostra que 7 deles têm uso mais acentuado na oralidade, apesar de também estarem presentes nas demais seções, fato este, que pode explicar por que figuram entre os dez mais da língua inglesa. *Grow up* apresentou um particularidade ao se destacar na seção *newspaper*, mas com frequência quase igual na seção *spoken*. Apenas dois PV do grupo dos 10 mais não obtiveram destaque maior na oralidade, a saber, *pick up* e *point out*.

Quanto ao *pick up*, ele apresentou um destaque muito mais acentuado na seção *fiction* (192,01), registro da escrita; mas também, a marca da seção *TV/movies* (114,37) demonstra que ele tem uso considerável na oralidade, isso sem considerar a marca da seção *spoken* (82,55), logo, fica claro que dados extraídos de pesquisas linguísticas na área da LC podem não apenas produzir conhecimento para satisfazer o gênio científico, mas também levantam informações que podem tornar os materiais de ensino representativos da realidade do uso linguístico. Assim, um material didático que traga uma aula destinada ao ensino da língua inglesa, direcionada para ensino-aprendizagem de PV em registros da ficção, ao optar por textos que tragam ocorrências de *pick up*, não só reflete a realidade linguística, como também pode preparar o aprendiz para desempenho satisfatório na escrita e também para a oralidade. Quanto à tendência de uso, no último período (2015-2019) houve uma queda em relação ao

equilíbrio apresentado desde 1990 até 2014. No que se refere às acepções dicionarizadas, o *Collins COBUILD Advanced Learner's English Dictionary* traz 15 acepções diferentes na entrada destinada à explicação dos sentidos expressos por *pick up*, assim percebe-se que novas pesquisas podem ser realizadas para levantar informações relevantes relacionadas à semântica associada aos registros em que o PV atua.

Point out é o único do grupo dos dez mais com destaque acentuado de uso no âmbito da *internet*, pois na seção *blog* apresenta uso de 100,72 *tokens* por milhão, e na seção *web*, aparece sua segunda área de destaque (83,68), e, mais surpreendente é o fato de apresentar o maior uso entre os dez na seção *academic* (78,30); apresentando ainda ocorrência maior que 70 aparições por milhão na seção *magazine*. Percebe-se então, que ao lidar com o ensino de PV, considerando *point out*, pode ser sensato associar seus aspectos léxico-gramaticais ao domínio de registros da escrita, e com o objetivo de interligar teoria e prática, pode ser relevante que o professor formador, no ensino superior, lance mão do PV, de modo consciente, não apenas na formação de futuros docentes, mas, acima de tudo para seu autodesenvolvimento, durante a escrita de textos em inglês para divulgação de suas descobertas, enquanto professor-pesquisador, que além de transmitir conhecimento, desenvolve papel ímpar na produção de novos conhecimentos. Vale apenas ressaltar, que é importante que se realizem novas pesquisas em torno de *point out*, pois sua tendência de uso tem demonstrado queda gradativa no decorrer das duas décadas amostradas no COCA.

Quanto aos que se destacam na escrita, é importante destacar que dois PV despontaram, por apresentarem números que indicam usos muito mais expressivos na seção *fiction*, e ocorrências relativamente baixas nas demais seções do córpus. Ao ler algum livro de ficção em língua inglesa é muito provável que o leitor terá contato com *look up*, e *look down*, pois o primeiro, de acordo com as amostras do córpus, apresenta ocorrência de 188,68 por milhão, enquanto o segundo, de 93,55 por milhão. Ambos apresentam uma ligeira queda em suas tendências de uso, de acordo com os gráficos proporcionados pelo COCA.

Há uma prática, entre desenvolvedores de material didático, de privilegiar textos da escrita, isto é, os chamados clássicos da literatura, ao definir o que compõe ou não determinado material, muitas vezes em detrimento da oralidade, denotando que uma seja falsamente superior a outra, assim, o que propomos é que ao invés de crenças ou práticas, abra-se espaço para uma práxis voltada para a aproximação da teoria e da prática, com a finalidade de munir os aprendizes da língua inglesa com conhecimentos que desencadearão o desenvolvimento de habilidades e competências significativas para vivência e convivência em espaços culturais e interculturais cada vez mais presentes em diversos contextos sociais.

Assim, as descobertas advindas de pesquisas da LC podem fornecer aporte significativo para o desenvolvimento da própria ciência e do ser humano concomitantemente, bastando apenas direcioná-las para os fins almejados.

Também é comum, no âmbito educacional/escolar, nos depararmos com textos da esfera jornalística, tais como, notícias, reportagens, editoriais, cartas do leitor e ao leitor etc., fato que abre caminho para que inúmeras pesquisas baseadas em *corpora* sejam realizadas para produzir informações capazes não só de descrever o comportamento léxico-gramatical de determinadas unidades linguísticas, mas acima de tudo, sejam aproveitadas para adequar as informações presentes no material didático ao uso da linguagem em seu curso natural. Nessa direção a título de exemplificação, ao consideramos os PV *set up* e *break down*, percebemos que seus usos são mais acentuados em textos das seções *magazine, newspaper, blog* e *web*.

Tais dados podem servir de insumo para a produção de materiais didáticos, ou até para o próprio professor ao planejar uma aula direcionada para que o aprendiz seja capaz de dominar vocabulário, estruturas e diferentes usos da linguagem em contextos reais, adequando assim, as práticas pedagógicas ao que recomendam documentos como BNCC, BNC-formação e Diretrizes do ensino superior e do Curso de Letras.

Pode-se relacionar essa questão à noção do que significa “saber uma língua” que, segundo Berber Sardinha (2010, p.296), “implica conhecer como dizer e escrever segundo convenções de variedades específicas da língua [...]; para isso, é preciso conhecer a léxico-gramática das escolhas necessárias e desejadas para aquela situação específica”, assim, fica claro que é importante não apenas ensinar as palavras e seus sentidos, como também, como, quando e onde empregá-las.

Quanto aos que se destacam na academia, além do *point out*, mencionado anteriormente, outros dois PV merecem destaque, pois apresentaram sua tendência de uso mais intensa na seção *academic*, a saber, *carry out* e *bring about*, o primeiro apresenta nesta seção uma marca de 62,08 ocorrências por milhão, enquanto o segundo, apesar de se destacar na seção, apresenta uma frequência mais modesta, igual a 22,92. Estas informações serviram de critério para a confecção de uma lista apontando as frequências mais altas.

Tabela 2 - Os 10 PV mais frequentes na seção *academic* do COCA, em 2020

#	PV	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
1	point out	100,72	83,68	6,74	75,67	35,48	72,23	53,72	78,30
2	carry out	20,60	32,81	6,08	22,37	8,48	20,38	23,85	62,08
3	go on	186,89	171,04	459,04	395,35	228,00	134,15	141,22	55,44

4	make up	57,58	60,15	42,85	40,83	46,97	56,81	53,78	40,73
5	sep up	64,67	65,22	46,82	53,88	44,89	73,97	72,44	38,74
6	take on	48,21	44,56	17,15	53,42	27,81	53,86	49,71	34,20
7	turn out	74,69	71,82	63,08	82,82	66,73	80,39	56,55	29,70
8	take on	48,21	44,56	17,15	53,42	27,81	53,86	49,71	34,20
9	bring about	10,56	12,22	1,63	7,94	2,54	7,95	5,77	22,92
10	take up	20,38	26,71	10,39	15,70	33,14	27,38	22,09	21,35

Fonte: Dados da pesquisa.

Nota 1: Os números em grifo são os mais altos para cada PV.

Nation (2001, p. 187 tradução nossa), ao tratar de listas de palavras acadêmicas, afirma que; “tipicamente, listas de vocabulário acadêmico incluem palavras que são comuns em textos acadêmicos e não tão comuns em outras áreas.¹⁰⁷”. Esse fato pode ser observado, por meio de consulta às tendências de uso, que apontam usos muito mais acentuados em textos acadêmicos do que em textos de outras esferas, evidenciando, dessa forma, que tal conhecimento pode contribuir para o ensino de ESP.

Ainda mencionando características de *carry oute bring about*, é relevante enfatizar que a tendência de uso do primeiro apresenta um leve aumento no período que compreende a última amostra do córpus (2015-2019), enquanto que o segundo apresenta uma acentuada tendência de que pode cair em desuso no futuro, se continuar a diminuir na mesma proporção.

Com relação à Tabela 2, as ocorrências em negrito, mostram em qual seção cada um se destaca. Os PV mais frequentes em textos acadêmicos, geralmente, têm uso mais acentuado em outros registros da linguagem e que a aprendizagem dos dez é relevante não apenas para satisfazer necessidades acadêmicas, pois realizam papel de destaque em textos de várias seções da amostra do córpus.

Como o presente trabalho lança um olhar especial sobre as necessidades do professor formador, pode-se dizer que o domínio dos dez PV pode ajudá-lo na tarefa de divulgar suas pesquisas por meio de emprego de vocabulário que realmente reflete o uso de falantes nativos da língua inglesa. De acordo com Nation (2003, p. 66), as palavras acadêmicas, apesar de desempenharem vários papéis na linguagem, figuram nessa lista, pois entre as funções que realizam na academia, estão “revisão de literatura, descrição de métodos de pesquisa, descrição de resultados etc.”

Quanto aos que se destacam na linguagem *online*, é importante levar em consideração que qualquer um dos dez mais frequentes da lista geral, figuraria nas primeiras posições da lista de destaques dos registros da *internet*. Vários dos PV que mostraram a marca mais

¹⁰⁷Typically, academic vocabulary lists include words[...] which are common in academic texts and not so common elsewhere.

acentuada de uso, seja na seção *blog* seja *na seção web*, são PV que figuram entre os menos frequentes da listagem dos 150. Embora tenham se destacado nas seções da linguagem virtual, suas ocorrências são bem inferiores às frequências apresentadas pelos primeiros da lista geral, por exemplo, *go on* que no *blog* possui ocorrências por milhão igual a 186,89; e na seção *web*, equivalente a 171,04. Como a intenção aqui não é listar os campeões de ocorrência em cada seção, e sim demonstrar características que envolvem apenas os que se destacaram nas seções da *internet*, em relação às demais, elencamos uma lista de dez exemplos, conforme Tabela 3.

Tabela 3 - PV com frequência de uso maior no registro internet

#	PV	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
1	end up	111,68	96,20	73,91	93,15	54,48	72,65	63,64	20,00
2	point out	100,72	83,68	6,74	75,67	35,48	72,23	53,72	78,30
3	make up	57,58	60,15	42,85	40,83	46,97	56,81	53,78	40,73
4	check out	56,21	43,24	44,19	19,49	16,93	33,03	18,79	4,8
5	keep up	30,04	26,18	17,17	13,64	22,79	20,94	19,74	7,40
6	put up	27,21	19,90	22,76	24,62	22,15	16,46	22,15	4,29
7	pass on	13,69	12,35	8,09	9,23	11,54	12,53	11,76	7,47
8	fill out	10,33	10,25	8,58	6,56	8,51	7,88	9,89	5,37
9	give out	7,10	6,16	5,78	6,52	7,00	4,31	5,25	1,53
10	give back	5,98	5,20	3,44	5,29	2,41	3,95	5,27	1,48

Fonte: Dados da pesquisa.

Destacamos, em primeiro lugar, que a preocupação com a representatividade do *córpus* e, conseqüentemente, com sua extensão e equilíbrio se justificam pela necessidade de conseguir capturar o maior número de palavras com frequências mais baixas. Se tivéssemos optado por um *córpus* pequeno, não abrangente e desequilibrado, provavelmente, algumas destas ocorrências não apareceriam.

Em segundo lugar, é relevante lembrar que a linguagem da *internet* tem se tornado cada vez mais importante e presente na vida de cada indivíduo. Desta forma, um programa de ensino moderno e adequado para as demandas do mundo contemporâneo deve contemplar essa variedade da linguagem em seu currículo de ensino tanto da língua materna, quanto de línguas estrangeiras. De acordo com McCarten (2007, p. 4, tradução nossa), “ferramentas de *córpus* podem nos dar informação sobre quão frequente uma palavra é em diferentes *corpora*, por isso nós podemos comparar a frequência do vocabulário na fala, jornais, textos acadêmicos e conversação¹⁰⁸”. As amostras linguísticas adicionadas na base de dados do COCA ampliam as

¹⁰⁸Corpus tools can give us information about how frequent a word is in different corpora, so we can compare the frequency of vocabulary in, say, newspapers, academic texts, and conversation.

possibilidades de compreensão da linguagem como um todo, permitindo comparar como determinado item lexical atua no ambiente virtual e em textos escritos, impressos ou falados.

Ao planejar o currículo de ensino, se as pessoas envolvidas optarem por seguir evidências oriundas de pesquisas de *corpora*, como as produzidas por esse estudo, proporcionarão uma experiência mais frutífera para os aprendizes. Optar por associar qualquer um dos quatro primeiros PV, listados na tabela 3, a textos da *internet*, pode abrir caminho para que o aprendiz tenha contato com vocabulário que ocorre na linguagem do mundo virtual. No caso de *end up* e *point out*, ainda, contribui para a apropriação de conhecimento que é significativo, tanto para fala do cotidiano, como para lidar com a linguagem de jornais e revistas.

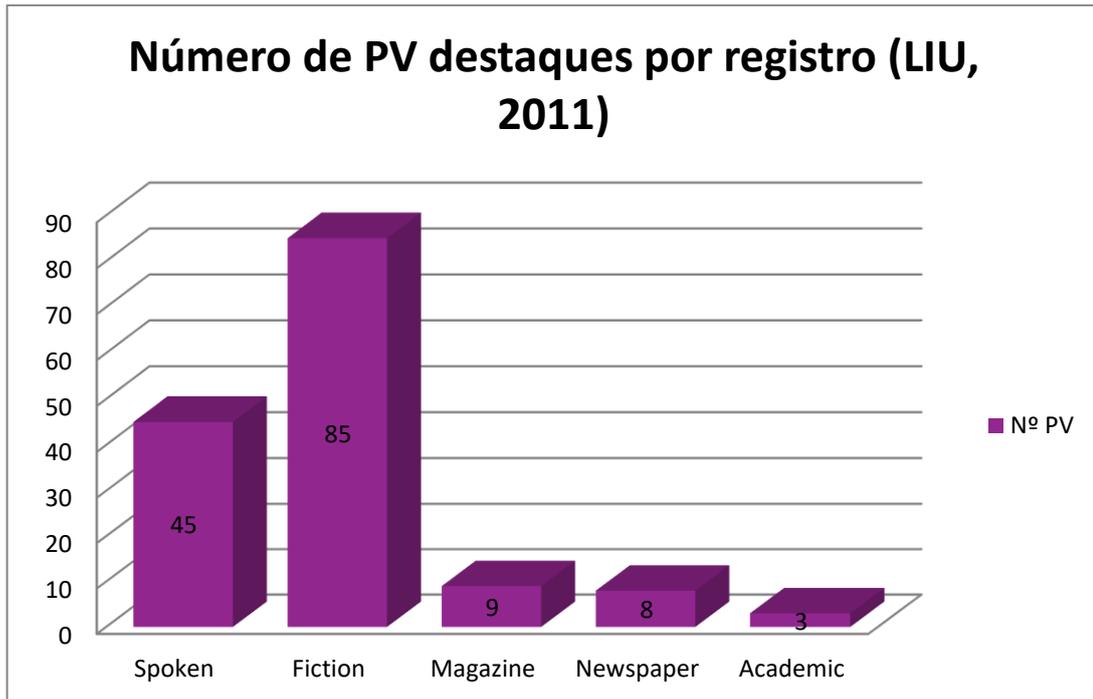
Na próxima seção, por meio da comparação dos dados da pesquisa de 2011 e os dados detalhados até aqui, pretendemos apontar situações em que pode ter havido mudanças neste intervalo de dez anos.

4.2 Comparação de uso de PV em 2011 e 2020 em termos de frequência e distribuição

Nesta seção, por meio da comparação entre as duas listas, analisamos o comportamento dos PV, no ano de 2011 e no ano de 2020, com o propósito de verificar se alguma mudança significativa ocorreu durante o período. Isto é, se houve continuidade, aumento ou diminuição do uso de determinado PV, em alguma das seções.

A Figura a seguir mostra quantos PV se destacaram em cada registro, de acordo com os dados da pesquisa realizada por Liu (2011).

Figura 7 - PV destaques por registro (LIU, 2011)



Fonte: Dados da pesquisa

Como pode-se verificar, a grande maioria dos 150 PV listados por Liu (2011) destacou-se nas seções *fiction* (85), demonstrando serem propensos ao uso em textos da escrita. O segundo grupo abarcou 45 PV que se destacaram na seção *spoken*, então, optamos por, em um primeiro momento, tratar das duas partes, e só então, apresentar os dados das outras três seções, pois apresentaram números bem menores. Para facilitar a exposição dos dados e, conseqüentemente, sua compreensão, agrupamos os dados referentes aos anos de 2011 e 2020, em uma única tabela, tornando assim, mais clara a visão de eventuais mudanças ou, até mesmo a continuidade da tendência de usos dos PV distribuídos por registros.

Tabela 4 - Os dez PV mais frequentes da variedade norte-americana em 2011 e 2020

	PV	2011					2020				
		Spok	Fict	Mag	News	Acad	Spok	Fict	Mag	News	Acad
1	go on	316,37	198,70	96,71	101,65	52,55	395,35	228,00	134,15	141,22	55,44
2	pick up	108,67	262,17	96,13	90,29	23,70	82,55	192,01	80,23	77,24	19,11
3	come back	251,75	163,84	50,15	69,15	11,92	232,61	157,90	51,34	68,47	11,85
4	come up	250,51	102,38	64,42	68,29	18,91	280,52	111,70	68,09	72,23	18,91
5	go back	190,98	154,92	56,15	65,03	19,74	232,61	157,90	51,34	68,47	11,85
6	find out	162,08	99,41	65,88	50,84	22,35	148,44	87,83	65,82	46,28	20,85
7	come out	163,22	90,99	45,35	50,32	11,48	244,70	134,14	72,93	80,33	19,63
8	go out	129,91	10,79	48,47	56,02	9,62	133,19	112,55	54,48	64,85	12,52
9	point out	79,13	38,72	76,70	62,05	90,72	75,67	35,48	72,23	53,72	78,30
10	grow up	89,25	61,39	75,95	97,38	22,57	94,05	65,47	78,05	96,32	21,24

Fonte: Dados da pesquisa

Uma vez que, ao se considerar apenas as seções *spoken* e *fiction*, nós estamos falando de exatamente, 130 dos 150 PV abordados nesta pesquisa, por uma questão de objetividade e praticidade, trataremos das tendências dos 10 primeiros da lista de Liu (2011), realizando assim um recorte viável para se alcançar os objetivos propostos.

Na seção *spoken*, sete dos dez foram mais salientes, a saber, *go on*, *come back*, *come up*, *go back*, *find out*, *come out* e *go out*. Destes, quatro apresentaram aumento em sua tendência de uso, dois apresentaram diminuição e um, por apresentar um aumento sutil, consideramos que se mantém estável.

Dentre os que se destacaram na seção *spoken*, *go on* chama a atenção por figurar como o mais frequente tanto nas listas de Davies e Gardner (2010, p. 218) e Liu (2011). Sua ocorrência naquele ano alcançou 316,37 por milhão e, dez anos depois, apresenta uma tendência de 395,35. Além de figurar como o mais frequente em 2020 (desconsiderando a frequência de *come on*, na seção *TV/movies*, por ainda não compor as amostras do COCA em 2011), sua frequência tende a crescer.

Dentre os PV, uma família de palavras composta por mais de 5.000 unidades dicionarizadas, parece ser uma boa ideia priorizar o ensino daqueles que realmente se destacam no dia a dia.

Além de *go on*, outros três apresentaram aumento na tendência de uso, *come up* passou de 250,51 para 280,52 em 2020; enquanto *go back* pulou de 190,98 para 232,61; e *come out* de 163,22 para 244,70. Quanto aos que apresentaram queda na seção *spoken*, destacam-se *come back*, que em 2011 apresentou ocorrência de 251,75; e em 2020, caiu para 232,21; e *find out*, cuja tendência saiu de 162,08 e desceu até 148,44. No caso de *go out*, a tendência de crescimento se apresentou de forma sutil, portanto, pode ser interpretada como uma estabilização no curso desta década.

Quanto à seção *fiction*, em que a maioria dos PV apresentou uso mais destacado, pode-se enfatizar o PV *pick up* cuja ocorrência em 2011 alcançou a marca de 262,17 por milhão e em 2020, segundo as amostras do córpus, sua ocorrência sofreu uma queda vertiginosa ao alcançar a marca de 192,01. Apesar da queda registrada, é um PV que pode figurar em materiais didáticos e carece do olhar científico da LC, em novas pesquisas, pois possui um campo semântico que se destaca entre os demais PV, pois segundo o dicionário LDOCE, possui 22 acepções diferentes, logo há um campo aberto para pesquisadores que se interessem em elucidar os usos de *pick up* e suas relações semânticas com os diversos registros da linguagem.

Com relação à seção *magazine*, que agrupa os registros que compõem o universo das revistas, nenhum dos dez primeiros se destacou nesse registro, assim, partimos em busca de *desert up*, o 11º na lista de 2011 e 16º em 2020, trata-se de um *phrasal verb* que se destaca neste meio, partiu de uma tendência em 2011, de 83,50 ocorrências por milhão e percorreu um caminho descendente até 2020, alcançando a marca de 73,97. Apesar de apresentar queda, é um PV que pode ser empregado para ensino de vocabulário com ênfase tanto em textos de revistas quanto em jornais, pois sua marca em 2020, de acordo com as amostras linguísticas do COCA, na seção *newspaper*, é de 72,44, fato que evidencia seu papel marcante em textos tanto da esfera jornalística, como também da esfera publicitária, presente em textos que podem ser encontrados em revistas e jornais. Possui um campo semântico com doze acepções diferentes, de acordo com o dicionário LDOCE, é importante ressaltar que apresentou tendência de diminuição de uso nas cinco seções do cópuz.

Na seção *newspaper*, pode-se destacar, dentre outros, *grow up* que figurou na seção com 97,38 ocorrências por milhão, segundo a pesquisa de 2011, já no presente estudo, figurou com 96,32; Apresentou uma queda sutil de uso em textos jornalísticos, o que não deprecia seu papel desempenhado na língua inglesa, pois continua apresentando uma tendência compatível a do início da década e, além disso, apresenta aumento em sua participação em três seções, a saber, *spoken*, *fiction* e *magazine*, logo, dominar este PV pressupõe habilidades comunicativas relacionadas à escrita e à oralidade evidenciando assim o quão importante é para o aprendiz da língua inglesa o conhecimento acerca do seu comportamento e de seu campo semântico que apresenta três acepções dicionarizadas. O comportamento de *grow up* pode ser melhor visualizado através da tabela a seguir.

Tabela 5 - Uso de *grow up* - 2011 e 2020

Ano	Spoken	Fiction	Magazine	Newspaper	Academic
2011	89,25	61,39	75,95	97,38	22,57
2020	94,05	65,47	78,05	96,32	21,24

Fonte: Dados da pesquisa

Por fim, no que se refere à seção *academic*, analisamos os três PV, cujas marcas mais altas de ocorrência compreendem aquelas relacionadas aos textos da esfera acadêmica, a saber, *point out*, *carry out*, *bring about*, os dois últimos, mesmo não figurando entre os onze mais frequentes, foram incluídos, devido à facilidade de alcançar a totalidade da seção.

Considerando *point out*, na pesquisa do início desta década, apresenta uma ocorrência de 90,72 por milhão, na seção *academic*, enquanto que em 2020, a ocorrência é de 78,30, evidenciando que pode haver uma tendência de queda em seu uso no meio acadêmico e ao comparar as ocorrências nas demais seções, percebe-se que essa tendência se manifesta em todas, porém em menor proporção. Apesar dessa diminuição, é notório que seu uso ainda é relevante seja na fala ou na escrita, nesta, o uso é mais acentuado em três seções, a do universo das revistas, das páginas do jornal e da esfera acadêmica. Logo, pode-se ensiná-lo e aprendê-lo associado aos textos da academia, dos jornais e das revistas em atividades de *reading* e *writing*, além de planejar atividades voltadas para o desenvolvimento de *listening* e *speaking*. Logo os dados extraídos trazem evidências que podem ser consideradas ao planejar um programa de ensino, um material didático, ou até mesmo ao planejar aulas específicas. A tabela 6 ilustra os dados expostos neste parágrafo.

Tabela 6 - Uso de point out - 2011 e 2020

Ano	Spoken	Fiction	Magazine	Newspaper	Academic
2011	79,13	38,72	76,70	62,05	90,72
2020	75,67	35,48	72,23	53,72	78,30

Fonte: Dados da pesquisa

Quanto ao *carry out*, ele não figura entre os dez mais frequentes, sua posição no ranking é a de número 36, e suas ocorrências de 62,25 e 62,08, em 2011 e 2020 respectivamente, sugerem estabilidade em seu uso em textos acadêmicos. Quanto ao campo semântico, duas acepções estão no dicionário LDOCE e uma delas se associa com '*research*' e '*survey*', que estão relacionadas ao contexto das pesquisas. Uma das contribuições da LC ao estudo da linguagem tem a ver com a descrição das associações mais frequentes, isto é, não se limita ao olhar sobre a ocorrência, focando também nos emparelhamentos, ou melhor, na co-ocorrência, fator fundamental na descrição da linguagem, uma vez que, a língua não é um sistema baseado em possibilidades, mas sim em probabilidades, logo, no contexto dos textos acadêmicos, *carry on*, *research* e *survey*, constantemente se atraem, formando assim colocados. Basear a produção de materiais didáticos na colocação contribui para que os textos reflitam uso da linguagem genuína e promove contato com traços linguísticos que facilitam a absorção de vocabulário e desenvolvimento da fluência. Os aspectos comportamentais de *carry out* estão expostos na tabela 7.

Tabela 7 - Uso de carry out - 2011 e 2020

Ano	Spoken	Fiction	Magazine	Newspaper	Academic
2011	26,48	12,13	19,92	23,83	62,25
2020	22,37	8,48	20,58	23,85	62,08

Fonte: Dados da pesquisa

No que tange ao uso de *bring about*, o número 107 da lista, percebe-se que o número de ocorrências por milhão tem caído durante o período dos últimos dez anos. Seu uso permanece mais acentuado no âmbito da academia, logo, apesar de figurar entre os mais frequentes na oralidade, sua aprendizagem pode auxiliar o professor formador na divulgação de suas pesquisas por meio da língua inglesa, pois além de estreitar a relação entre teoria e prática, promover conhecimento relevante para si mesmo e para os discentes, torna seus textos mais próximos do uso linguístico do nativo, transpondo barreiras culturais e promovendo a interculturalidade, esta uma demanda contemporânea da globalização. A tabela 8 facilita a compreensão dos traços linguísticos que circundam o uso de *bring about*.

Tabela 8 - Uso de bring about - 2011 e 2020

Ano	Spoken	Fiction	Magazine	Newspaper	Academic
2011	10,72	2,80	8,67	6,51	27,44
2020	7,94	2,54	7,95	5,77	22,92

Fonte: Dados da pesquisa

4.3 Tendência de uso dos PV em contextos formais e informais

Nesta seção, para dar sequência na pesquisa, primeiramente, dividimos os registros do COCA em dois grupos, partindo do princípio de que os usos mais formais podem ser encontrados nas seções *fiction*, *magazine*, *news* e *academic*, e que os menos formais podem ser encontrados nas seções *blog*, *web*, *TV/movies* e *spoken*. Em seguida, elencamos 3 PV em cada uma das seções, considerando as frequências mais altas.

Nessa direção, 8 PV serão analisados no primeiro grupo, a saber, *Go on*, *look up*, *pick up*, *Come in*, *turn out*, *grow up*, *point out* e *carry out*. Enquanto que, no que se refere ao segundo grupo, os menos formais, analisaremos as ocorrências de 6 PV, a saber, *go on*, *end up*, *come out*, *come on*, *get out* e *come up*.

Ao passo que as análises aconteceram, consultas nos dicionários foram realizadas, para verificar se há alguma informação, mencionando se tal uso é mais recorrente em contextos tidos como mais formais ou menos formais.

4.3.1 Contextos mais formais

No que tange ao primeiro grupo, partimos por “*go on*”, a fim de encontrar contextos mais formais da seção *Fiction*. Essa seção é atualmente composta por 119,505,292 palavras, distribuídas em textos de histórias curtas e peças de revistas literárias, revistas infantis, revistas populares, primeiros capítulos de primeiras edições de livros de 1990 até os dias atuais e roteiros de filmes. Optamos por realizar buscas na subseção *Movies/documentary*, onde encontramos um filme intitulado *Manufacturing Consent: Noam Chomsky and the Media*, lançado nos Estados Unidos, em 1992, sua classificação de gênero é: documentário, biografia e guerra. Um filme sobre o notável linguista/dissidente político americano e seu aviso sobre o papel da mídia corporativa na propaganda moderna. O gráfico a seguir, mostra como o PV se comporta nas diferentes subseções do registro *Fiction*.

Sobre as ocorrências de “*go on*” no filme, selecionamos os seguintes contextos de uso do PV:

Quadro 6 – Contextos de *go on* extraídos do filme *Manufacturing Consent: Noam Chomsky and the Media*

Léxico-gramática	Contextos
Infinitivo	The reason that the atrocities can go on is because nobody knows about them. If anyone knew about them, there'd be protests and pressure to stop them.
Gerúndio/presente perfeito contínuo	The extensive thinking that's been going on for a long, long period, about the necessity for finding ways to marginalise and control the public in a democratic society.
Gerúndio/passado contínuo	Literally dropped to zero. All this was going on at exactly the same time as the great protest of outrage over Cambodia.
Gerúndio presente contínuo	You read about your neighbours, see what's going on in the district, and things like that.
Presente simples	They're supposed to vote. They're supposed to play some role in the way economic and political and cultural life goes on .

Fonte: Dados da pesquisa

Todos os contextos extraídos desta amostra exemplificam a acepção “acontecer” pode-se perceber o nível de formalidade através do assunto tratado e do sujeito emissor das falas, um linguista muito famoso e que também se dedica a causas políticas. Nesse sentido, fica evidente que as estruturas léxico-gramaticais associadas aos contextos de produção da mensagem podem contribuir não apenas para o entendimento de traços linguísticos, mas

também para o desenvolvimento da competência crítico-reflexiva do aprendiz, uma vez que entram em contato não apenas com estruturas gramaticais e vocabulário da língua, como também oportuniza reflexões importantes sobre questões político-econômicas que cercam o aprendiz. O fato evidencia que os dados produzidos pela LC podem ser realmente relevantes para o desenvolvimento pleno do ser humano, beneficiando assim não apenas a área de Ensino de Língua Inglesa, mas, acima de tudo a educação. No que se relaciona com o tratamento dado pelos dicionários, as duas amostras analisadas, o Collins COBUILD e o LDOCE, apenas mostram quais contextos são informais, abordaremos a questão quando estudarmos os PV e contextos menos formais.

O próximo PV analisado será “*pick up*”. Sua disposição nas subseções da seção *fiction* está ilustrada na figura 8 a seguir:

Figura 8 – Comportamento de *pick up* nas subseções da ficção

Gen (Book)	Gen (Jrn)	SciFi/Fant	Juvenile	Movies	Fan Fiction
6800	7818	4088	737	3072	219
33.4	44.8	26.2	3.2	9.2	1.5
203.43	174.53	156.31	227.83	334.83	142.35

Fonte: COCA (2020)

Para analisar as ocorrências de “*Pick up*” em contextos mais formais da seção *fiction*, selecionamos alguns usos extraídos da subseção *TV: documentary*. A série escolhida foi *Nature*, lançada nos Estados Unidos em 1982, com 33 episódios, classificada nos gêneros documentário e família, explora várias questões e assuntos globais da vida selvagem. O gráfico na sequência mostra o comportamento das estruturas léxico-gramaticais no âmbito geral.

Figura 9 – Panorama geral das estruturas léxico-gramaticais de *pick up*

HELP		ALL FORMS (SAMPLE): 100 200 500	FREQ	
1	<input type="checkbox"/>	PICK UP	36871	<div style="width: 100%;"></div>
2	<input type="checkbox"/>	PICKED UP	31537	<div style="width: 85%;"></div>
3	<input type="checkbox"/>	PICKING UP	11445	<div style="width: 31%;"></div>
4	<input type="checkbox"/>	PICKS UP	6991	<div style="width: 19%;"></div>
5	<input type="checkbox"/>	PICK UPS	37	<div style="width: 0.1%;"></div>
		TOTAL	86881	

Fonte: COCA (2020)

Percebe-se que a forma base (*present/imperative/infinitive*) e a forma acrescida de “ED” (*past simple/past participle*) são muito mais recorrentes, uma provável explicação para o fenômeno é que estas formas atuam em várias construções gramaticais distintas, atuando

assim em vários contextos, por exemplo, o passado simples é muito recorrente em gêneros narrativos, enquanto a forma base pode atuar no presente simples, no imperativo e no infinitivo. A seguir, apresentamos um quadro com alguns contextos de “*pick up*” selecionados na série *Nature*.

Quadro 6 – Contextos de *pick up* extraídos da série *Nature*

Léxico-gramática	Contextos
Presente contínuo	Man: Wolverines don't hibernate, and so they're out all winter long and they're feeding on animals that are killed by avalanches and falls. So they're picking up the pieces of other animals that didn't make it through the winter.
Presente simples	Narrator: They pick up a signal from one of the collars.

Fonte: Dados da pesquisa

Nos contextos selecionados da série, temos respectivamente amostras das acepções pegar/coletar e detectar um sinal. É importante ressaltar que nos dois contextos há a ocorrência de contrações, traço linguístico mais recorrente em textos falados, logo, percebe-se que, embora se trate de um documentário, a oralidade permite essa mistura, demonstrando que a linguagem é dinâmica e heterogênea. No seu curso normal, principalmente na fala que é marcada pela espontaneidade, dificilmente usaremos uma linguagem puramente formal, independente do contexto, exceto quando as mensagens são emitidas por pessoas douras e que querem deixar esse traço em destaque.

Pick up é um PV cujo campo semântico é muito amplo, o dicionário LDOCE, por exemplo, traz 22 usos diferentes para esse PV, apenas uma acepção é marcada pelo uso informal, nesse dicionário.

Para iniciar a descrição dos comportamentos do PV *look up*, apresentamos um gráfico com suas frequências em cada subseção de *fiction*.

Figura 10– Tendências de usos de *look up* nas subseções de *fiction*

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
FIC:Gen (Book)	5570	33.4	166.63	
FIC:Gen (Jrnl)	7027	44.8	156.87	
FIC:SciFi/Fant	4793	26.2	183.27	
FIC:Juvenile	662	3.2	204.64	
FIC:Movies	3876	9.2	422.47	
FIC:Fan Fiction	417	1.5	271.05	

Fonte: COCA (2020)

De acordo com o gráfico, percebe-se que a seção da ficção é subdividida em 6 gêneros, e seu destaque está em *Movies* com uma fpm igual a 422, 47. De acordo com a frequência geral, “*look up*” é mais recorrente apenas na seção *Fiction*, pois sua fpm é de 188,68, muito maior do que aquelas apresentadas nas demais seções, pois não superam a marca das 40 fpm. Para buscar contextos mais formais de sua tendência de uso na seção *Fiction*, coletamos contextos

de uma subseção chamada “*Harvard Review*”. As amostras foram extraídas de várias publicações, com autores diferentes. Os textos são narrativos, com presença de verbos no passado, personagens, enredo e um narrador em 1ª pessoa. Os contextos extraídos desta amostra linguística estão no quadro que segue, vale destacar que apenas a segunda amostra é um PV, as demais são falsos PV, pois ao analisar os sentidos produzidos, percebemos que se tratava de um verbo (olhar) e a partícula adverbial indicava o lugar para onde se dirigia o olhar (para cima).

Quadro 8 – Contextos de *look up* extraídos da seção *Harvard Review*

Léxico-gramática	Contextos
Passado simples	Luis looked up at two thousand feet of bluing rock. He said, " I don't know, man. "
Gerúndio	All she had to do was lay out the plan, and Sammie took over, looking up the address and phone number, calling several mornings in a row to make sure Mrs. Farber would be home. *
Passado simples	I looked up and down, up and down, until the upper left of my margin was filled with her features. Then she laughed when the professor said something, and I saw that her teeth were yellow.
Passado simples (negativa)	The room was crowded with the tables and chairs that had been pushed against the walls to make room and we stood girls on one side boys on the other, you didn't look up because the ceiling was black with fag smoke and you didn't look down because the floor was sticky from old drink so all you had to look at was each other, which for some reason was all of a sudden unbearable, and it took several songs before anyone crossed over to the other side for a turn and then things got going a bit, but I always danced with my brother, I was having none of that business, so me and Cathal did a sort of swirly dance
Passado simples	Maybe Luis was thinking about that as he looked up at the granite cliff. Maybe he was already regretting what he had said in Camp 4.

Fonte: Dados da pesquisa

*Ocorrência de PV verdadeiro

Como se tratam de textos narrativos, a tendência é que ocorram mais estruturas léxico-gramaticais no passado. Apenas a segunda amostra é um PV, ao consultar as acepções dos PV no dicionário, encontramos 3 acepções distintas, a saber, “melhorar”, “buscar informação” e “visitar”. Uma vez que a maiorias dos contextos é de falso PV, uma nova busca foi iniciada, insistimos nas buscas na subseção destinada a periódicos, mais duas ocorrências foram encontradas, então, destinamos nossos esforços para a subseção “*movies*” que resultaram em mais duas ocorrências verdadeiras, uma no filme *Hardware* – acepção de “melhorar”; e uma outra ocorrência no filme *Jacob’s Ladder* – acepção de “procurar”. Os contextos de uso estão demonstrados no quadro a seguir.

Quadro 9 – Contextos de *look up* extraídos da subseção *journal e movies*

Léxico-gramática	Contextos
Futuro	Io says she will look up what to do on the Internet. (<i>Crazyhorse/Rainforest</i> , 2019)
Infinitivo	The dusk outside is a murky blue, and inside, the living room furniture darkens into

strange densities. He can't remember how to enter Privacy Mode; the transition itself gets key-indexed along with the protected memories, so he has to **look up** the instructions each time. (*The Magazine of Fantasy and Science Fiction/Lacuna heights*)

Presente contínuo	Things are looking up . Just one more reason to celebrate on this great day, when we honor the birth of that very special man... and... yet he was not a man(melhorar - <i>Hardware</i> - 1990)
Presente simples	The locks set. JACOB begins instantly rifling through a desk drawer. He comes up with a frayed address book and looks up a number. (procurar – <i>Jacob's Ladder</i> – 1990)

Fonte: Dados da pesquisa.

Na primeira ocorrência, a do futuro, a acepção empregada de *look up* é procurar, fazer uma busca na internet. A segunda ocorrência, também traz a mesma acepção, pois é preciso buscar as instruções. A quarta, mais uma vez, traz a acepção de procurar, mas agora um número.

Na terceira ocorrência, há uma amostra de uma acepção diferente, pois “*look up*” também pode significar “melhorar”, logo as coisas estão melhorando. Os dicionários não trazem informação acerca dos usos formais e informais, provavelmente, pode-se entender que suas acepções trabalham normalmente tanto em mensagens formais quanto informais.

A seção *Magazine*, próximo alvo da investigação é atualmente composta por 127,352,014 palavras. Quase 100 revistas de diversos domínios como *news, health, home and gardening, women, financial, religion, sports etc.* Alguns exemplos de revistas populares que constam no cópua são: *Time, Men's Health, Good Housekeeping, Cosmopolitan, Fortune, Christian Certury* e *Sports Illustrated*.

Parece uma boa ideia buscar pelos contextos de uso nas revistas de saúde e de economia, a fim de encontrar bons exemplos de usos mais formais.

Figura 11 – *Come in* nas subseções de *Magazines*

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
MAG:News/Opin	2030	27.7	73.35	
MAG:Finacial	785	7.2	109.30	
MAG:Sci/Tech	1507	18.6	80.94	
MAG:Soc/Arts	607	9.3	65.10	
MAG:Religion	293	5.6	52.76	
MAG:Sports	1716	14.1	121.52	
MAG:Entertain	693	7.7	89.79	
MAG:Home/Health	1807	20.7	87.30	
MAG:Afric-Amer	361	4.3	84.65	
MAG:Children	282	2.4	116.50	
MAG:Women/Men	737	8.5	86.57	

Fonte: COCA 2020.

O gráfico deixa claro que o PV em questão é muito mais recorrente em três subseções, destacando-se na seção *sports* com uma fpm igual a 121,52. Alguns contextos foram selecionados e destacados no quadro 10:

Quadro 10 – Contextos de *come in* extraídos da seção *Magazines*

Léxico-gramática	Contextos
Presente simples	The owls come in two colors, pale gray and reddish-brown. Until recently, selection favored the pale gray shade; those owls were more likely to survive in the snowy Finland landscape. (<i>Science news</i> – 2014)
Possibilidade	Amazon recently warned that its expectations for sales this holiday season could come in below Wall Street forecasts. (<i>Fortune</i> – 2018)
Infinitivo	We have also had amazing success with colleges and universities and intern programs and teaming up with other organizations to get volunteers to come in and run part of our programs. (<i>Total Health</i>)
Possibilidade	Increased domestic production could come in handy, even if it is hardly enough to replace the shortfall of Iraqi and Kuwaiti oil. (<i>Newsweek</i> – 1990)

Fonte: Dados da pesquisa.

A primeira ocorrência, à primeira vista, pode parecer um exemplo de falso PV, pois é comum dizer que um objeto vem em determinada cor, por exemplo, bicicletas vêm em várias cores e modelos, porém parece mais apropriado traduzir *come in*, no caso das corrujas, como elas ocorrem em duas cores, portanto não nos parece coerente, tomar “*come in*” como um falso PV nesse contexto. Vale ressaltar que um dos aspectos que diferem PV dos verbos preposicionados é justamente o fato de novos sentidos surgirem da associação dos elementos composicionais.

Quanto à segunda ocorrência, não há menção à essa acepção nos dicionários pesquisados neste trabalho, logo fica claro que os dados produzidos aqui podem enriquecer os dicionários, enfatizando assim, a contribuição que a LC e os *corpora* podem proporcionar aos trabalhos dos lexicógrafos.

A terceira amostra exemplifica a acepção “juntar”, “fazer parte”, “se envolver”. Os dicionários não fazem menção a usos formais ou informais, uma provável explicação é que o PV tende a ocorrer nos dois contextos.

A quarta e última ocorrência traz uma nova acepção não destacada nas entradas do PV “*come in*” nos dicionários analisados, logo, essa informação também é útil para enriquecer as próximas edições dos dicionários, a nova acepção é “ser útil” ou (*come in*) *handy*. Os dicionários lidam com essa acepção na entrada “*handy*”, porém não destacam o emparelhamento como um *chunk* formado pela associação de um PV mais um adjetivo.

O próximo PV analisado é “*turn out*”. Segundo os dicionários LDOCE e Collins COBUILD, há sete acepções em seu campo semântico, logo se trata de um PV com certa densidade em sua carga de aprendizagem e suas tendências de uso são bem amplas como podemos perceber ao analisar os diferentes contextos de uso demonstrados da figura 12:

Figura 12 – Turn out na seção Magazine

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
MAG:News/Opin	2852	27.7	103.05	
MAG:Financial	656	7.2	91.34	
MAG:Sci/Tech	1667	18.6	89.53	
MAG:Soc/Arts	637	9.3	68.32	
MAG:Religion	307	5.6	55.28	
MAG:Sports	829	14.1	58.71	
MAG:Entertain	824	7.7	106.77	
MAG:Home/Health	1263	20.7	61.02	
MAG:Afric-Amer	185	4.3	43.38	
MAG:Children	149	2.4	61.56	
MAG:Women/Men	795	8.5	93.39	

Fonte: COCA (2020)

Embora os textos das subseções “*Financial*” e “*Science/Technology*” não sejam os contextos mais frequentes de atuação de “*turn out*”, parecem ser boas fontes para buscar contextos mais formais. Tais contextos estão explicitados no quadro 11.

Quadro 11 – Contextos de *turn out* extraídos da seção *Magazines*

Léxico-gramática	Contextos
Possibilidade	They also know that the field remains littered with good ideas that didn't pan out. # " It could turn out to be a big bust in the next several years, " Johnson admits. " But I think it's a good lead. " # PHOTO: This picture shows messenger RNA for a complement protein found in brain tissue taken from a patient with Alzheimer's disease.[<i>Science News</i> - 1992]
Passado simples	However, when the carnivorous snails arrived, they turned out to prefer the taste of native Achatinellae to that of the African pests they were intended to eradicate. # " Cannibal snails are devouring the last of the Oahu tree snail species, " [<i>Science News</i> - 1992]
Conditional	"We were not really aware of how rich these regions would turn out to be, "says Gatley "There are lots and lots of young stars in there. We had wanted to look at these areas since the 1960s, but we needed to see past the dust to find out what was going on." [<i>Science News</i> - 1991]
Presente simples	Endothelial cells turn out to play key roles in maintaining the circulatory system. " They not only can sense flow, but they can do something about it, " says John A. Frangos, a chemical engineer at Pennsylvania State University in University Park. [<i>Science News</i> - 1991]
Presente simples	It turns out that even though the cells were identical, the electrical devices that powered them were not. " Nobody ever thought power supplies would affect the plasma, but that really shows how touchy the system is, " says Roberts. [<i>Science News</i> - 1991]

Fonte: Dados da pesquisa

Todos os contextos estudados de “*turn out*” podem ser traduzidos como “resultar”, isto é, acontecer de um determinado modo, então, vale destacar que o emparelhamento da segunda amostra “*turn out to prefer*” causou certa dúvida no processo de tradução para o português, desse modo, acreditamos que poderia ser uma boa ideia incluí-lo nos dicionários, pois nem o LDOCE ou o Collins COBUILD tratam dessa estrutura léxico-gramatical.

A próxima seção que receberá o foco de nossas lentes de investigação é a *Newspaper* que atualmente conta com 122,959,393 palavras no cópulo. São dez jornais norte-americanos, por exemplo, *USA Today*, *New York Times*, *Atlanta Journal Constitution*, *San Francisco Chronicle*. Em muitos casos, há uma mistura de seções de jornais como notícias locais, esportes, opinião, mercado financeiro etc. Nesta seção, serão pesquisados “*grow up*” e “*come in*”.

Figura 13 – *Grow up* na seção *News*

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
NEWS:Misc	3031	31.7	95.49	
NEWS:News_Intl	277	6.3	43.81	
NEWS:News_Natl	973	11.5	84.43	
NEWS:News_Local	1604	13.8	116.03	
NEWS:Money	371	8.7	42.52	
NEWS:Life	2949	20.3	145.59	
NEWS:Sports	2137	19.2	111.57	
NEWS:Editorial	384	4.8	80.04	

Fonte: (COCA 2020)

De acordo com as distribuições dos textos, independente das frequências apresentadas, a fim de listar contextos mais formais, decidiu-se por utilizar como fontes de pesquisa, as subseções *Editorial* e *Money*. Assim, o quadro a seguir traz os contextos listados para a descrição das tendências de uso de “*Grow up*”.

Quadro 12 – Contextos de *grow up* extraídos da seção *Newspaper*

Léxico-gramática	Contextos
Passado simples	Lee grew up in the shade of her sister, who was nearly four years older and was her father's favorite and an accomplished equestrian. They lived in commodious Manhattan apartments and at estates on Long Island and in McLean, Va., and Newport, R.I., where they learned to sail on Narragansett Bay. [<i>New York Times</i> – 2019]
Passado simples	Supreme Court Justice Thurgood Marshall, who grew up in Baltimore, was denied admission to the University of Maryland in 1930 because of his race. He attended Howard University instead, and would go on to help overturn the legalized segregation of schools in Brown v. Board of Education. [<i>Baltimore Sun</i> – 2019]
Passado contínuo	Mark Donen acknowledged Saturday that his father wasn't around much when he was growing up . And "he could be pretty fierce on a movie set. He wasn't particularly interested in collaboration -- unless it was with someone he respected very much [<i>Chicago Tribune</i> – 2019]
Passado simples	Matthews, who grew up in Northern Virginia, said that he wanted to apologize to "the many I have hurt through my actions " and that there was no justification for them. [<i>Washington Post</i> – 2019]

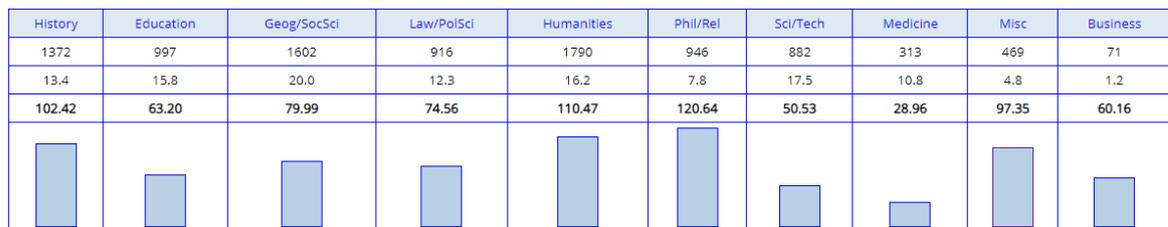
Fonte: Dados da pesquisa

Todos as amostras selecionadas estão no passado, uma vez que se trata de textos narrativos da esfera jornalística, há uma tendência de que muitas ocorrências remetam ao passado, foram 118 páginas com 11725 concordâncias do PV disponibilizadas pelo COCA, há várias ocorrências do gerúndio e do infinitivo, mas atreladas a outros verbos geralmente no

passado. “*Grow up*” é um PV com apenas 3 acepções dicionarizadas, a saber, “se tornar adulto”, “cresça/amadureça!” e “desenvolver”. A segunda acepção aparece com uma nota nos dicionários indicando que essa acepção é predominante da fala, pois se trata de uma frase exclamativa e com sentidos pejorativos, já que estabelece a ideia de que a pessoa é imatura, apesar da idade.

A seção *Academic*, objeto das próximas análises, atualmente, possui uma amostra linguística de 120,988,348 palavras, composta por quase cem jornais de divulgação de pesquisas acadêmicas abordando áreas como filosofia, psicologia, religião, história do mundo, educação, tecnologia etc. O gráfico seguinte mostra a composição da seção *Academic* e as frequências de “*Point out*” o primeiro PV analisado nessa seção:

Figura – 14 As tendências de uso de “*point out*” na seção “*Academic*”



Fonte: COCA (2020)

Esta é uma seção marcada pela formalidade e pelo rigor dos textos acadêmicos, logo não será problema encontrar amostras de uso do PV em contextos mais formais. Ele se mostra mais recorrente na subseção *Philosophy/Religion*, por esse motivo, vamos buscar contextos de uso nessa subseção. Ao acionar busca na seção *Academic*, o COCA disponibilizou uma lista de 10 páginas e 946 concordâncias de *point out*. A tabela a seguir traz alguns dos contextos encontrados, apenas a título de exemplificação, já que não há espaço e nem tempo hábil para estudar todas as concordâncias:

Quadro 13 – Contextos de *point out* extraídos da seção *Academic*

Léxico-gramática	Contextos
Presente simples	Still in the context of this seminar, Heidegger points out once more how the difference between the human and the animal lies in this can-say, which directly acquires an ontological significance. [<i>Philosophy Today</i> – 2017]
Passado simples	If, as I already pointed out , for Heidegger " living means: being in an animal kind of way " then the animal " does not exist but merely lives " (Heidegger 1995: 210) [<i>Philosophy Today</i> – 2017]
Presente perfeito	If, as I already pointed out , for Heidegger " living means: being in an animal kind of way " then the animal " does not exist but merely lives " (Heidegger 1995: 210) [<i>Philosophy Today</i> – 2017]
Presente perfeito	as Derrida has pointed out , Heidegger has probably not really overcome the horizon of metaphysics, and has hence remained ensnared in the meshes of his own theoretical fabric, thus continuing to spiritualistically believe that the world is either " always the world of spirit " or it is not even world. [<i>Philosophy Today</i> – 2017]

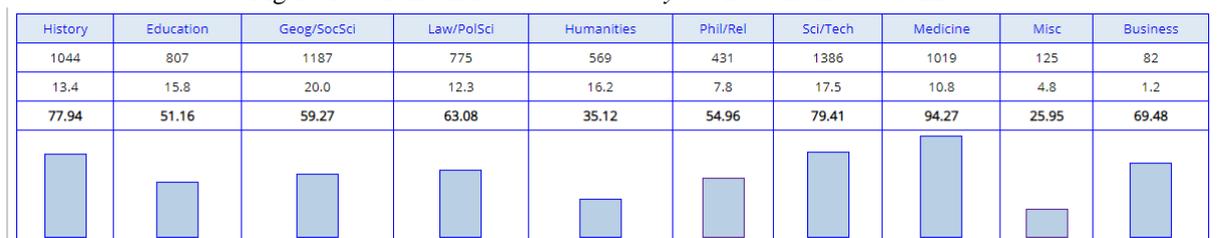
Passado Perfeito	Moreover, as Michel Haar had already pointed out with a particularly brilliant sharpness of mind, the persistence of such a spiritualist perspective in Heidegger's thought on animality would be essentially due to the fact that Heidegger can not really think of the animal because he has in mind from the very beginning the ontological unicity of man. Indeed, by separating human Dasein from animality, he would only perform an actual " exorcism " of man's animal nature, his main preoccupation being to demystify the link between human being and nature. In other words, [<i>Philosophy Today</i> – 2017]
Presente simples	Dalia Judovitz rightly points out that, " Descartes's deliberate use of pictorial analogies is not intended for the exploration of the visible, but rather its foreclosure " (Judovitz 1993: 84). [<i>Philosophy Today</i> – 2017]

Fonte: Dados da pesquisa

Levando em consideração os contextos extraídos, percebe-se que se trata de um verbo *dicendi*, isto é, um verbo de elocução, utilizado para demarcar a fala de alguém, no caso, várias citações de diferentes vozes ao longo do texto, é uma estratégia argumentativa frequentemente utilizada em textos do âmbito acadêmico. Vários tempos verbais são empregados, demonstrando, que além da assimilação da escrita acadêmica, pode-se desenvolver a competência linguística por meio da aprendizagem de diferentes estruturas léxico-gramaticais. Duas acepções são dicionarizadas de acordo com o LDOCE e Collins COBUILD, mas não há demarcação sobre preferência de uso em contextos formais ou informais. Seria uma boa estratégia, se os lexicógrafos inserissem um contexto de uso de citações e demarcasse que se trata de uma tendência dos textos acadêmicos.

Na sequência analisaremos as tendências de uso de “*carry out*”

Figura 15 – Tendências de uso de *carry out* no contexto acadêmico



Fonte: COCA (2020)

Ao analisarmos o gráfico, salta aos olhos a maior ocorrência de “*carry out*” em contextos da medicina, logo realizaremos buscas em jornais da subseção “*Medicine*”. Os contextos selecionados estão disponibilizados no quadro que segue:

Quadro 14 – Contextos de “*carry out*” extraídos da seção *Academic*

Léxico-gramática	Contextos
To be + past participle	Reconstruction of sections was carried out with GPU-based scanner software (NRecon). The grayscale was based on the Hounsfield unit, and the validated calcium standards were scanned as its density reference. [<i>Indian Journal of Orthopaedics</i> -

To be + past participle	2019] All animal procedures were carried out according to the Guide for the Care and Use of Laboratory Animals and were approved by the Committee of Experimental Animal Sciences of our institute. [<i>Indian Journal of Orthopaedics</i> - 2019]
To be + past participle	A search was carried out using the following search string: " (((Primary meningococcal arthritis) OR Monarthritis Neisseria meningitidis) OR Primary Neisseria arthritis) or Neisseria monoarthritis meningitidis) or Isolated meningococcal arthritis. " We found a total of 79 items (Fig. 1). The title and background were reviewed. [<i>Bulletin of the NYU Hospital for Joint Diseases</i> - 2019]
To be + past participle	Interviews were carried out using a mixture of open pre-determined and follow-up questions. The interviews allowed participants to talk about their role and the types of duties that they undertook, their experience of the transition of public health from NHS to LA, and their experience and involvement in providing evidence for, or making a decision around, commissioning a public health service or making a policy change. [<i>Health Research Policy and Systems</i> - 2019]
To be + past participle	The search was carried out against UniProt database (November 2016 release) containing human proteins as well as E. coli beta-gal. The result (. group) file thus generated served as the spectral ion library. [<i>Journal of Translational Medicine</i> – 2019]
To be + past participle	The PCR reaction was carried out on a Veriti Thermal Cycler (Life Technologies) and ran as follow: an initial denaturation step (10 min at 95C), 10 cycles (95C for 15 seconds, 60C 30 seconds, 72C 1 min), 2 cycles (95C for 15 seconds, 80C 30 seconds, 60C 30 seconds, 72C 1 min), 8 cycles (95C for 15 seconds, 60C 30 seconds, 72C 1 min), 2 cycles (95C for 15 seconds, 80C 30 seconds, 60C 30 [<i>PLoS Neglected Tropical Diseases</i> - 2019]

Fonte: Dados da pesquisa

Ao analisarmos os contextos extraídos dos diversos jornais da área da medicina, pode-se perceber que é comum o emparelhamento do verbo “to be” no passado com o PV no participípio. Esta estrutura é comum em partes do texto que explicam os procedimentos metodológicos empregados, a sua aceção nesse contexto é “realizar”, não há marcações nos dicionários LDOCE e Collins COBUILD, nesse sentido, é importante ressaltar a necessidade de se incluir contextos extraídos de procedimentos metodológicos no verbete do PV e incluir a informação referente ao uso formal.

Na sequência, abordamos os PV nos contextos menos formais.

4.3.2 Contextos menos formais

Com o intuito de listar contextos de usos menos formais, vamos verificar as tendências de uso de seis PV (*come on, go on, get out, come up, end up e come out*), em quatro seções do corpús (*TV/Movies, Spoken, Blog, Web*), cujas circunstâncias de produção das mensagens podem ser marcadas por textos produzidos em contextos mais espontâneos e livres de rigor de formalidade. Nessa perspectiva, iniciamos os estudos por *TV/Movies*, esta seção conta com 128,013,334 palavras de textos de programas da TV americana e legendas de filmes, estas legendas podem ser tão informais ou mais informais que os dados da fala em contextos de uso real do cotidiano.

O primeiro PV analisado é “*come on*”. Segundo o dicionário *Collins COBUILD*, há 6 acepções diferentes atribuídas a “*come on*”, a saber, 1) uma expressão de encorajamento (recorrente na fala), 2) uma expressão utilizada para apressar alguém (recorrente na fala), 3) início dos sintomas de doenças, 4) fazer progresso, 5) iniciar o funcionamento de uma máquina, e 6) chegada de novas estações. Enquanto que o dicionário LDOCE atribui 10 acepções ao mesmo PV, a saber, 1.a) apressar alguém, 1.b) encorajar alguém, 1.c) dizer que algo é mentira ou errado, 1.d) provocar alguém, 2) convidar alguém para entrar de modo amigável, 3) iniciar o trabalho de uma máquina, 4) começar a sentir-se doente, 5) iniciar um programa de TV ou rádio, 6) começar a chover ou nevar, 7) alcançar um estágio, 8) melhorar ou fazer progresso, 9) encontrar alguém por acaso, 10) dizer que alguém é sexualmente atrativo. Merece destaque o fato de que todas as entradas do grupo 1 são apontadas como recorrentes na fala pelo LDOCE, e que a acepção 10 é a mais informal.

Antes de iniciar as análises, é importante mostrar a família de palavras relacionadas a *come on*, de acordo com as amostras do *Corpus of Contemporary American English* e elencadas na figura 15, apresentada na sequência.

Figura 16 – Lista lematizada das ocorrências de *come on*

HELP		ALL FORMS (SAMPLE): 100 200 500	FREQ	
1	<input type="checkbox"/>	COME ON	210229	
2	<input type="checkbox"/>	CAME ON	6638	
3	<input type="checkbox"/>	COMING ON	4107	
4	<input type="checkbox"/>	COMES ON	2385	
5	<input type="checkbox"/>	COMETH ON	5	
6	<input type="checkbox"/>	COMIN ON	3	
7	<input type="checkbox"/>	COMEING ON	1	
TOTAL			223368	

Fonte: COCA (2020)

A lematização de “*come on*” apresentou dados surpreendentes, tais como, “*cometh on*” e “*comeing on*”. Quanto à frequência, vale destacar que “*come on*” apresenta uma tendência de uso muito maior do que as demais formas. Uma possível explicação é que a forma básica atua no presente, no imperativo, no infinitivo e ainda cumpre o papel de passado participio, englobando assim, os tempos perfeitos.

“*Cometh on*” não aparece nos dicionários, seu uso é restrito a textos bíblicos, entretanto, quando comparamos algumas passagens bíblicas citadas nos contextos do COCA com as passagens em uma Bíblia mais atual, descobrimos que pode ser uma forma antiga que caiu em desuso, pois não aparece mais. É importante destacar que seus contextos de uso estão limitados às seções *Blog* e *Web*, as postagens são do ano de 2012.

Vale destacar também a ocorrência de “*comin on*”, com 3 ocorrências no corpus, para ser mais específico, na seção *fiction*. Os dicionários não tratam destas formas do gerúndio com

a omissão do “g” final. Seria interessante que os lexicógrafos atentassem para esse fenômeno e acrescentassem essa possibilidade de ocorrência aos dicionários para facilitar a jornada de futuros aprendizes da língua inglesa.

Para abordar o PV na seção *TV/Movies*, primeiramente, verificamos as frequências nas subseções e, então escolhemos uma fonte aparentemente adequada para explorar o PV em contextos tidos como informais. A figura 16 emitida pelo COCA facilita a compreensão da organização da seção.

Figura 17 – Frequências de “come on” nas subseções de *TV/Movies*

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT
MovAction	22516	9.2	2,444.44	
MovAdult	367	0.2	1,531.02	
MovAdv	6073	2.4	2,510.02	
MovAnim	3477	2.3	1,544.37	
MovBlog	4041	2.3	1,733.56	
MovComedy	32495	18.4	1,765.86	
MovCrime	6622	3.8	1,720.72	
MovDocum	3013	6.9	437.11	
MovDrama	20319	11.6	1,751.93	
MovFam	815	0.3	2,349.20	
MovFantasy	932	0.4	2,315.30	
MovHorror	8834	3.3	2,716.65	
MovMusic	182	0.2	851.76	
MovMyst	553	0.4	1,531.54	
MovRomance	240	0.2	990.29	
MovSci-Fi	224	0.2	1,139.60	
MovShort	416	0.3	1,655.46	
MovThrill	451	0.4	1,156.55	
MovN/A	1554	0.7	2,077.85	
TVAction	13038	12.5	1,046.67	
TVAdv	3365	2.8	1,201.03	
TVAnim	4671	4.7	996.03	
TVComedy	23308	19.1	1,222.06	
TVCrime	7901	9.4	839.43	
TVDocum	280	1.4	202.97	
TVDrama	13271	13.0	1,023.92	
TVGame	757	0.7	1,120.35	
TVHorror	118	0.1	990.81	
TVReality	271	0.4	755.53	
TVSci-Fi	118	0.1	1,308.29	
TMisc	298	0.3	1,176.36	

Fonte: COCA (2020)

De acordo com o gráfico, é fácil perceber que ele é mais recorrente na subseção *Horror*, então decidimos colher amostras de usos do PV em contextos de filmes de terror. Os contextos estão explicitados no quadro 15.

Quadro 15 – Contextos de “come on” extraídos da seção “*TV/Movies*”

Léxico-gramática	Contextos
Imperativo	I'll take care of him. - I'll take you down right here. - Go, go, go! Come on! You heifer-butt, I'm gonna fuck you in hell! [<i>Leatherface: Texas Chainsaw Massacre III</i> - 1990]
Imperativo	Maybe I can bring the bad guys out in the light. You coming? All right, you stay here. Don't move. I'll be right back. Hold up. Keep it. Thanks. Eee! Nice neighborhood.

Imperativo	Damn. - Come on . - Aah! Aah! [<i>Leatherface: Texas Chainsaw Massacre III</i> - 1990] Red alert. Red alert. Blue sector. Come on . Kim. Red alert. Red alert. This is it. Oh, no. My security pass. [<i>Metamorphosis: The Alien Factor</i> – 1990]
Imperativo	Norman, what are we doing here? Come on! I tricked you, I had to. I hated it, but I didn't know any other way. I love you so much. I want us to have a baby so much. [<i>Psycho IV: The Beginning</i> – 1990]
Imperativo	hat's cute. You all go have a good time. All right, come on now, Joel, come on back. Hey, stop playing. Bring my girl back over here, man. [<i>Def by Temptation</i> – 1990]
Imperativo	Come here. Yeah. Come on . Come on . Go get it. Go get it. That's it. Stay out of those flowers. [<i>Buried Alive</i> – 1990]

Fonte: Dados da pesquisa

As buscas resultaram em 89 páginas com 8833 linhas de concordâncias de “*come on*”. As concordâncias deixaram claro que esse PV tem seu uso impulsionado pelas acepções relacionadas ao encorajamento e ao apressar o interlocutor. São frases imperativas e exclamativas. Quando comparamos sua frequência na seção *TV/Movies* com a *Spoken*, percebemos que a fpm nos filmes é incrivelmente superior a da fala do cotidiano. Pode-se depreender desse fato que apesar da linguagem utilizada nos filmes ser uma representação da fala do mundo real, sua frequência aponta para sua natureza artificial.

O próximo PV investigado na seção *TV/Movies* é “*go on*”. De acordo com o dicionário LDOCE, há 15 acepções para o PV, sendo 4 acepções marcadas como usos de contextos informais. Estas correspondem às entradas: 12) falar muito, 13) criticar, 14) desenvolver ou fazer progresso, e finalmente 15) ter algo suficiente para o momento. O dicionário Collins COBUILD apresenta 11 acepções para “*go on*”, destas 2 são destacadas como usos de contextos informais, a saber, 8) continuar a falar da mesma coisa, 9) expressão de encorajamento. O quadro 16, na sequência, traz algumas amostras de contextos extraídos da seção *TV/Movies*.

Quadro 16 – Contextos de “*go on*” extraídos da seção “*TV/Movies*”

Léxico-gramática	Contextos
Gerúndio	In the war castle scene, there's a lot going on under the surface. The most important thing about that scene is understanding the lay of the land and understanding what their intentions are, what their plans are, what their expectations are, so we know what we're looking at. [<i>Game of Thrones</i> – 2019]
Presente contínuo	What made you leave? The war. The war ended when you were a kid. Well, the war is still going on for my family. Uh... my two older brothers died at Gettysburg. [<i>Warrior</i> – 2019]
Presente contínuo	Jace, we have to do it. Do what? Clary, what's going on ? I spoke to Catarina. She won't summon Lilith. Too dangerous. [<i>Shadow hunters</i> – 2019]
Imperativo	Uh, you were comparing me to a prize stud. Go on . You have fun with your new best buddy? Hey, it's your fault your dad likes me. [<i>The Big Bang Theory</i> – 2019]
Presente simples	I know all about your tall girl website. It goes on our credit card. It's Dr. Pemberton. Kripke says he plagiarized his thesis, and that he can prove it.
Presente contínuo	Move! Up. Hey. Get up. What the hell's going on ? Early release for good behavior. What about the bombs in our necks? [<i>Arrow</i> – 2019]

Fonte: Dados da pesquisa

A busca resultou em 588 páginas com um total de 58790 linhas de concordâncias. O gerúndio e imperativos são visivelmente mais recorrentes. As formas do gerúndio são amostras da aceção acontecer, enquanto a forma do imperativo é utilizada para encorajar ou incentivar, podendo ser traduzida para o português como “Vai” ou “Continua”. O último contexto é mais informal, pois a construção vem acompanhada de um palavrão, “hell”, em português: “inferno”. A pergunta é traduzida para o português como “Que inferno está acontecendo?” Ela aparece 1.593 vezes na seção *TV/Movies*. O gráfico a seguir mostra de modo simplificado as frequências da pergunta “*What the hell is going on?*” nas seções do COCA.

Figura 18 – Frequência de “*What the hell is going on?*”

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	1923	53	43	1593	49	147	26	12	0	243	304	333	312	279	356
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	1.94	0.41	0.35	12.44	0.39	1.24	0.21	0.10	0.00	1.75	2.06	2.27	2.15	1.92	2.46
SEE ALL SUB-SECTIONS AT ONCE															

Fonte: COCA (2020)

Vale lembrar que o gráfico não apresenta os números da forma contraída, isto é, “*What the hell's going on?*”, a contração apresenta números bem menores, mas permanece mais recorrente na seção *TV/Movies*, como se pode ver no gráfico a seguir.

Figura 19 – Frequência da forma contraída “*What the hell's going on?*”

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	1090	2	4	941	24	110	6	3	0	216	207	158	187	148	168
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	1.10	0.02	0.03	7.35	0.19	0.93	0.05	0.02	0.00	1.55	1.40	1.08	1.29	1.02	1.16
SEE ALL SUB-SECTIONS AT ONCE															

Fonte: COCA (2020)

Ao compararmos as figuras 18 e 19, podemos perceber que a pergunta se destaca nas seções *TV/Movies* e na *Fiction*, mais uma vez, as ocorrências da fala do cotidiano são muito inferiores da tendência de uso nos filmes, então há um campo aberto para novas pesquisas, para atestar se há evidências de que a linguagem da ficção reflete modelos reais do cotidiano, porém com frequências muito superiores que podem demonstrar que a linguagem da ficção é artificial.

Finalizaremos as investigações na seção *TV/Movies* com os estudos da tendência de uso de “*get out*”. O dicionário LDOCE apresenta 7 acepções diferentes, a saber, 1) cair fora, 2) escapar, 3) ajudar alguém a escapar, 4) pegar algo, 5) deixar escapar informação secreta, 6) produzir, e 7) dizer algo. Importante ressaltar que o LDOCE não faz marcações de uso informais para este PV. Quanto ao Collins COBUILD, o dicionário apresenta apenas 4 acepções para “*get out*”, 1) deixar um lugar, 2) sair em busca de uma vida melhor, 3) deixar o trabalho, e 4) deixar informação ser conhecida. O COBUILD também não faz marcações de usos informais. O gráfico a seguir, deixa claro que “*get out*” se destaca na seção *TV/Movies*.

Figura 20– Frequências de “*get out*” nas seções do COCA

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS AT ONCE)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
BLOG	8,627	128.6	67.08	
WEB-GENL	7,800	124.3	62.77	
TV/MOVIES	44,761	128.1	349.49	
SPOKEN	16,354	126.1	129.65	
FICTION	16,546	118.3	139.84	
MAGAZINE	6,040	126.1	47.90	
NEWSPAPER	6,341	121.7	52.09	
ACADEMIC	1,112	119.8	9.28	

Fonte: COCA (2020)

Além de se destacar da seção de filmes, o PV também é recorrente nas seções *Fiction* e *Spoken*, os números muito elevados na seção *TV/Movies* também apontam para a possibilidade de que as tendências de uso na linguagem cinematográfica podem refletir a linguagem do cotidiano, porém as frequências ressaltam o caráter artificial das ocorrências. O quadro a seguir, traz contextos de uso de “*get out*”.

Quadro 17 – Contextos de “*get out*” extraídos da seção “*TV/Movies*”

Léxico-gramática	Contextos
Imperativo	Let's get out of here, lads. Move! Thank you. Excuse me. Excuse me, miss. I just wanted to say thank you. For what you done. [<i>Carnival Row</i> – 2019]
Imperativos	Hey, chink. Chinks, go home! Chinks, go home! Get out of my way. Christ, these people stink. [<i>Warrior</i> – 2019]
Presente contínuo	I don't want any trouble. I'm getting out of here today. Oh, I know how you feel. I broke out of prison today. [<i>Arrested Development</i> – 2019]
Passado Simples	But I warn you, I just got out of a weird relationship, and I might complain about my ex a lot. [<i>The Big Bang Theory</i> – 2019]
Present perfect	She wrecked my family. But we could've done it the right way. She would've gotten out eventually. You don't know that. At least now we know she'll never hurt anyone ever again. I get it. But killing her? [<i>Shadowhunters</i> – 2019]
Modal verb – necessidade	Must be after sunrise. We gotta get out of here now. Careful. Careful. Just because we don't hear anything doesn't mean Heidi didn't leave more vamps to finish us off. [<i>Shadowhunters</i> – 2019]

Fonte: Dados da pesquisa

Todos os seis contextos selecionados exemplificam a acepção “deixar o local, sair, escapar”, a quarta amostra é a única que não se relaciona a um lugar, mas sim a um relacionamento amoroso. Todos os contextos são bem casuais, porém o último contexto é marcado por um índice de informalidade maior devido à estrutura léxico-gramatical construída ao redor de “*gotta*”. De acordo com o dicionário LDOCE, trata-se de uma estrutura recorrente na fala e marcada pelo tom de informalidade, inclusive as pessoas acreditam que seu uso é agramatical. O dicionário *Collins COBUILD* diverge do anterior, pois destaca que “*gotta*” é recorrente na escrita, mas ambos concordam no contexto marcado pela informalidade. A figura a seguir mostra a frequência do emparelhamento “*gotta get out of here*”.

Figura 21 – ocorrências de “*gotta get out of here*”

SECTION	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD	1990-94	1995-99	2000-04	2005-09	2010-14	2015-19
FREQ	855	1	1	811	3	35	2	2	0	149	162	147	157	120	118
WORDS (M)	993	128.6	124.3	128.1	126.1	118.3	126.1	121.7	119.8	139.1	147.8	146.6	144.9	145.3	144.7
PER MIL	0.86	0.01	0.01	6.33	0.02	0.30	0.02	0.02	0.00	1.07	1.10	1.00	1.08	0.83	0.82
SEE ALL SUB-SECTIONS AT ONCE															

Fonte: COCA (2020)

O emparelhamento é muito mais recorrente na seção *TV/Movies*, e supera em mais de 20 vezes sua frequência na seção *Fiction*, sua segunda maior frequência apontada pelo COCA, mais uma vez, salta aos olhos do linguista que a linguagem do cinema reflete usos da linguagem do dia a dia, entretanto, o número de ocorrências pode evidenciar sua artificialidade. Se o emparelhamento fosse utilizado em um programa de ensino de língua inglesa, certamente poderia aparecer em textos de diálogos de filmes, pois assim, seu uso estaria coerentemente ilustrado nesse contexto. Quanto ao PV “*get out*”, seria uma excelente escolha utilizá-lo para o ensino do imperativo, empregando contextos extraídos de filmes.

Agora, apresentamos os estudos realizados na seção *Spoken*, que conta com 127,396,916 palavras, são transcrições de mais de 150 programas de rádio e TV, tais como *All Things Considered*, *Newshour*, *Good Morning America*, *Today Show*, *60 Minutes*, *Hannity and Colmes*, *Jerry Springer etc.* O COCA ainda destaca a autenticidade e naturalidade das amostras que compõem esta seção.

Os estudos iniciaram-se pelo PV “*come up*” que segundo o dicionário LDOCE, apresenta 11 acepções, a saber, 1) aproximar-se para falar, 2) viajar ao norte, 3) surgir um assunto, 4) surgir um problema, 5) surgir uma oportunidade, 6) lidar com um caso judicial, 7)

acontecer em breve, 8) surgimento do sol ou da lua, 9) brotar, 10) vomitar, 11) pronto para consumo. Apenas a última possui uma marcação destacando que se trata de uso da linguagem falada. O *Collins COBUILD* destaca apenas 7 acepções, porém percebe-se que as acepções relacionadas à “surgir” estão incluídas em uma só entrada. Importante destacar que o *Collins* não traz nenhuma marcação de contextos de fala ou informalidade para o PV. Alguns contextos extraídos do COCA estão explicitados no quadro a seguir.

Quadro 18 – Contextos de “*come up*” extraídos da seção “*Spoken*”

Léxico-gramática	Contextos
Gerúndio	I think you have a football game coming up in your state. I can't remember who is playing. [CNN Newsroom- 2019]
Futuro (going to, would e will)	Bill Barr has right now two congressional hearings scheduled this week, tomorrow and the next day. If the next day, one, actually happens, this is something that's certainly going to come up at those. I'm not sure it would come up before we reported it or not. It will certainly come up now. [CNN: Erin Burnett OutFront - 2019]
Gerúndio	Always good to have you here. Coming up next, a look at this morning's other headlines including why the brothers involved in the alleged Jussie -- Jussie Smollett hoax are now suing two of the Empire actor's attorneys. [CBS News: CBS This Morning- 2019]
Presente simples	We'll stir all of that together. It comes up looking like this? Exactly. And then you just get some nice bibb lettuce or butter lettuce. [NBC News: Today – 2019]
Gerúndio	Coming up , the panel and the next steps on the road to possible impeachment. But first, how small country under siege from Russia became a threat to Donalnd Trump's presidency. [NBC News: Meet the Press – 2019]
Presente perfeito	Anthony, today you've come up with a couple of delicious ways to use peanut butter from our sponsor Jif. What do we got? [NBC News: Today – 2019]

Fonte: Dados da pesquisa

Ao atentarmos para os contextos do gerúndio, pode-se perceber que em programas de televisão, “*coming up*” é utilizados para dizer ao telespectador o que vem na sequência, poderíamos, então traduzir o PV como “a seguir”. A forma do presente perfeito “*have come up*” exemplifica a acepção “produzir”, no último contexto, Anthony produziu novas formas deliciosas de empregar a manteiga de amendoim, neste contexto, o apresentador (a) faz uso do PV para persuadir o telespectador a consumir o produto do patrocinador. Aqui, percebe-se que dominar a palavra requer muito mais do que apenas conhecimento gramatical e semântico, pois, acima de tudo é preciso ter conhecimento das finalidades dos usos da língua, seu uso funcional, para então organizar as palavras de modo que os objetivos sociais sejam alcançados. Nesse sentido, a língua é tomada em uma perspectiva de Letramentos, pois demanda o desenvolvimento de práticas sociais situadas. É importante salientar o papel social que as pesquisas da LC realiza, pois pode contribuir para a compreensão de aspectos linguísticos e extra-linguísticos capacitando os aprendizes da língua inglesa a empregarem a língua de modo consciente e coerente, realizando assim seus objetivos sociais, ao passo que

lançam mão da linguagem. A próxima figura ilustra os usos do PV em diversas emissoras e o quanto é mais recorrente em três delas.

Figura 22 – Tendências de uso de “come up” na seção *Spoken*

ABC	NBC	CBS	CNN	FOX	MSNBC	PBS	NPR	Indep
4333	4691	4746	7800	5097	354	1762	5358	1243
20.1	10.5	17.6	26.3	12.2	0.8	11.3	22.7	4.5
215.78	445.32	269.24	296.08	419.29	429.12	155.70	235.66	275.28

Fonte: COCA (2020)

Na sequência, o PV “come out” foi o foco dos estudos. De acordo com o dicionário LDOCE, há 14 acepções ligadas aos usos de “come out”, a saber, 1) remover, 2) vir à tona, 3) revelar, 4) lançar, 5) resultar, 6) produzir sentidos, 7) terminar, 8) ser perceptível, 9) posicionar-se favorável/contra, 10) aparecer, 11) desabrochar, 12) sair do armário (assumir a homossexualidade), 13) recusar a trabalhar (greve), e 14) debutar. Apenas a acepção realcionada às greves é marcada como uso informal no LDOCE. O *Collins COBUILD*, por sua vez traz apenas 8 acepções e assiná-la “go on strike” como um uso restrito ao contexto britânico, demonstrando a preocupação com os aspectos interculturais do uso da língua inglesa. Apesar de não ser um dos objetivos desta pesquisa, vale ressaltar que a LC tem aberto caminho para vários estudos com corpora com a finalidade de comparar usos em variedades distintas de uma mesma língua, caso queiram se desafiar nessa empreitada, sugerimos a utilização do *Corpus of Contemporary American English* (COCA) e do *British National Corpus*(BNC) como fontes de pesquisas para comparação de usos nas variedades norte-americana e britânica da língua inglesa. Alguns contextos de “come out” estão exemplificados no quadro a seguir.

Quadro 19 – Contextos de “come out” extraídos da seção “Spoken”

Léxico-gramática	Contextos
Passado simples	Three of his top rivals came out for outlining the kind of health insurance that 200 million Americans rely on. They are -- they are on a -- in a race to the left on immigration. [CBS News: <i>Face the Nation</i> – 2019]
Passado perfeito	Most of all the Democrat competitors to Kamala Harris have come out in support of her. Well, I don't think that Republicans are afraid of having an argument about busing in the 1970s. Let's recall, busing was (INAUDIBLE) with white Americans – BRENNAN). [CBS News: <i>Face the Nation</i> – 2019]
Presente simples	Second, he said, I do not believe it is appropriate for people to lie about contacts with Russians. When the report comes out , what the American people will learn was that there was plenty of reason to be concerned about the nature of those contacts and the possibility of what those contacts might have meant. [Fox News: <i>Fox News Sunday</i> – 2019]
Gerúndio	That movie coming out next year, this time next year. I love that. There was so much to choose from this morning. It was one of those days, there was a lot of great

Fonte: Dados da pesquisa

O COCA listou 30.865 linhas de concordância, em 309 páginas, fontes ricas e preciosas para a realização de inúmeras descrições lingüísticas tanto quantitativas quanto qualitativas.

O primeiro contexto está ligado à questões políticas vivenciadas pelos norte-americanos, pode-se perceber, então, que o conhecimento deste PV pode contribuir para a formação plena do aprendiz, uma vez que além dos conhecimentos léxico-gramaticais em jogo, existe a possibilidade de reflexão-crítica acerca de sua situação no mundo, elevando o aprendiz a uma condição de sujeito, pois ao refletir sobre como atua no mundo, pode-se formar um ser pensante, questionador e reflexivo, que não apenas se adapta ao mundo, mas que também percebe as contradições e busca transformar o mundo, tornando-se assim agente, movido pela criatividade e autonomia, as quais podem ser potencializadas no processo de ensino-aprendizagem da linguagem.

O segundo contexto, também está ligado a questões políticas, mas lembrando fatos de um passado distante, ao abordar a questão dos ônibus e da segregação racial. Nesse sentido, nota-se que a aprendizagem de uma nova língua requer também conhecimento histórico-cultural acerca dos falantes da língua-alvo, portanto é necessário lançar mão de um olhar intercultural que pressupõe conhecer o outro com suas similitudes e contradições, e ao entrar em contato com essa cultura nova, esse outro diferente, abre-se caminho para uma reflexão do “eu”, não para iniciar um processo de aculturação, mas, para que se tome consciência da pluralidade cultural existente dentro uma mesma sociedade, ou dentro de outras comunidades com língua-cultura completamente diferentes.

Na mesma direção, o terceiro contexto trata das relações turbulentas entre norte-americanos e russos, logo, fica claro que o ensino de línguas baseado em textos autênticos, oferece material não apenas para reflexão de estruturas léxico-gramaticais, pois textos produzidos por falantes nativos, no curso real do uso da linguagem, são fontes de conhecimento útil não apenas para se compreender a língua, mas acima de tudo, a cultura destes falantes, com seus costumes, crenças e atitudes.

O quarto contexto selecionado pela pesquisa mostra o PV empregado para expressar a subjetividade do falante, portanto, fica evidente que, nos textos oralizados na TV norte-americana, a tendência de uso do PV aponta para a função referencial, pois é utilizado para veicular notícias sobre acontecimentos de relevância social, mas também é utilizado para expressar os próprios juízos de valor, isto é, a subjetividade.

Portanto, ao verificar os usos dos PV nos contextos da seção *spoken*, percebe-se que cumprem inúmeras funções comunicativas e que esse conhecimento é muito importante para aprendizes da língua inglesa, pois revelam traços da língua-cultura. Essa visão intercultural proporcionada pelas análises *cross-register* produz descrições do uso da língua que se incluídas em gramáticas, dicionários e materiais didáticos, podem enriquecê-los e torná-los mais úteis aos seus usuários. Dicionários, por exemplo, são utilizados não apenas para auxiliar no ensino da língua, mas também para auxiliar viajantes a se comunicarem eficazmente em contextos reais de uso da língua, nesse sentido, os dados autênticos sobre o uso da língua inglesa nos dicionários podem ajudar substancialmente na comunicação no mundo globalizado, principalmente quando desenvolvidos para contextos específicos.

Para encerrar as análises dos contextos mais informais, trataremos do PV “*end up*” nos contextos da linguagem da internet. Vivemos em um mundo dividido pelo contexto real e pelo contexto virtual. Essa divisão, de fato, não sintetizou os usos da linguagem, mas sim ampliou as possibilidades de uso da língua, pois no mundo contemporâneo é preciso saber como se comunicar em contextos tanto do mundo real quanto do mundo virtual. Portanto, exercer uma cidadania plena requer o conhecimento de como “surfear essa nova onda” chamada internet. A figura a seguir traz as frequências de “*end up*” em cada seção, é notório que se trata de um PV de destaque em contextos menos formais, visto que suas frequências maiores estão nas seções marcadas pela informalidade, a saber, *Blog*, *Web*, *TV/Movies* e *Spoken*:

Figura 23 – Frequências de “*end up*” nas seções do COCA

SECTION (CLICK FOR SUB-SECTIONS) (SEE ALL SECTIONS)	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
BLOG	14,364	128,6	111,68	
WEB-GENL	11,953	124,3	96,20	
TV/MOVIES	9,466	128,1	73,91	
SPOKEN	11,749	126,1	93,15	
FICTION	6,446	118,3	54,48	
MAGAZINE	9,162	126,1	72,66	
NEWSPAPER	7,748	121,7	63,64	
ACADEMIC	2,396	119,8	20,00	

Fonte: COCA (2020)

Atualmente, a seção *Blog* conta com 125,496,215 palavras compiladas em uma subseção do *Corpus of Global Web-Based English* (GLOWBE). Naquela época, o *Google* permitia buscas restritas a *blogs*, logo quase todos os textos são de fato *blogs*. Alguns contextos extraídos do cópulus estão representados no quadro a seguir.

Quadro 20 – Contextos de “*end up*” extraídos da seção “*Blog*”

Léxico-gramática	Contextos
Presente simples	Unfortunately I don't have that privilege as our area games for the most

Passado simples	part end up being on the same day same time. [...]Of course I would then have to have a web player for each stream on my site etc. Then when the games end up being over, I would then have to go through the process of moving the archive [http://forum.sourcefabric.org/discussion/14505/multiple-live-broadcast-at-the-same-time/p1]
Presente simples	While short term this worked out nicely, but long term this ended up a big mess. What ended up happening was that in 6 months my team was rewriting these features and spending the a substantial amount of our time trying to keep their low-quality code up and running; falling even farther behind on creating the new features needed by the business. Not to mention being a major drag on my team who would love to work on new features, but were stuck in sustainment hell. # It wasn't productive for us, for the sales team, and especially not for our company. [http://katemats.com/engineer-whispering/]
Passado simples	12 Monkeys Following a man sent back in time to avert a global catastrophe, who ends up tossed in an asylum, this one is a fantastic movie that works well beyond the premise. [http://blastr.com/2012/10/x-of-the-most-convoluted.php]
Futuro simples	So, in my forgiving state, I actually ended up loving that film quite a bit (CERTAINLY better than Nemeshit). # But... now... I admit, I'm having a really tough time with the sequel news so far. The name alone makes me cringe. [http://www.aintitcool.com/node/58847]
Presente contínuo	I give a 20 percent chance that Sandy's storm surge will end up flooding a portion of the New York City subway system, " he said. # As he noted, the city had a narrow escape from Hurricane Irene, and in some ways this storm poses greater risks. [http://thelede.blogs.nytimes.com/2012/10/27/coastal-surge-forecast-from-hurricane-sandy-prompts-evacuations/]
	(although I have no doubt that we could find hundreds of billions of dollars of 'waste' that's ending up in the pockets of the 1%). [http://www.dailykos.com/story/2012/11/09/1159760/-Something-important-you-need-to-read-about-the-Fiscal-Cliff-before-the-Republicans-Tantrum-starts]

Fonte: Dados da pesquisa

De acordo com o dicionário LDOCE, “*end up*” possui apenas uma acepção, a saber, “acabar em alguma situação, estado ou lugar”, não há marcações nesse dicionário sobre contextos de uso informais ou se ocorre na fala ou na escrita. Segundo o *Collins COBUILD*, há duas acepções relacionadas a esse PV, pois é possível “acabar em um lugar” ou “acabar em determinada situação”, também não traz nenhuma marcação de uso destacado em contextos informais, de fala ou escrita.

A busca resultou em 14364 linhas de concordância lematizadas distribuídas em 144 páginas. Na verdade, como se pode observar nos contextos extraídos do corpus, trata-se de um PV com um amplo campo de atuação, apesar de ter um campo semântico limitado à apenas uma acepção.

Por exemplo, no primeiro contexto, o assunto do *blog* é um jogo, portanto, jogadores lançam mão deste PV para expressar como os jogos acabam e em que situação eles próprios acabaram. Nesse *blog*, claramente há um contexto marcado pela informalidade e que inclusive pode trazer gírias utilizadas pela comunidade de jogadores.

O segundo contexto traz amostras de usos do PV em um contexto um pouco mais formal, pois o autor expressa questões relacionadas à engenharia, destaca a sua equipe e a

companhia, logo é possível perceber que não está ligado a contextos de entretenimento, mas sim ao contexto profissional.

O terceiro contexto traz pistas de se tratar de um *blog* de resenhas de filme, pois o autor expressa suas percepções e juízos de valor acerca do filme. O quarto exemplo também representa uma resenha de filme.

O quinto e o sexto contextos destacam assuntos mais sérios, a tratarem respectivamente de um furacão e de questões financeiras. Os usos da linguagem são culturalmente influenciados, por exemplo, nos Estados Unidos há o costume de tratar tempestades e furacões com nomes de mulheres, Sandy e Irene são o nomes das tempestades citadas no *blog*. Na sequência, o mesmo PV foi analisado na seção *Web*.

A seção *Web* é hoje uma compilação de textos com 129,899,426 palavras de textos gerais da *Web*, há em sua compilação textos que na verdade são *blogs*, pois em 2012, ano em que foram coletados, o *Goggle* não contava com a opção “não-*blogs*” em suas ferramentas de busca. Diferentemente das demais seções, *Blog* e *Web* não permitem comparações diacrônicas, pois são como uma fotografia do ano da compilação do corpus. Na sequência, destacamos alguns contextos selecionados para discussão.

Quadro 21 – Contextos de “*end up*” extraídos da seção “*Web*”

Léxico-gramática		Contextos
Passado simples + verbo no gerúndio		Quinn: We were arguing about the Cheerios and I ended up calling her names and rubbing it in her face that Brittany was still technically with Artie # [http://www.fanfiction.net/s/7360490/1/Sorry-Seems-To-Be-The-Hardest-Word]
Could + infinitive		Any one of us could potentially end up in a position such as Tony's, so it is an issue relevant to us all. # Individuals should be allowed to make choices, that is the point of a democracy and in this instance, the quality of life is very poor and the suffering unforgivable. Can we have a bit of humanity please? [http://www.change.org/petitions/tony-nicklinson-s-right-to-die-change-the-law]
Presente adjetivo	simples +	But ultimately, if you have a confrontation between to people and no witnesses and one guy ends up dead and the live guy ends up with injuries, chances are the live guy is going to get off. [http://reason.com/blog/2012/03/26/stand-your-ground-may-have-nothing-to-do]
Presente verbo no gerúndio	simples +	As I've said in the past, no one has ever set forth a good measurement for what makes a Manager of the Year, so it usually ends up with a " who did the most with the least " analysis. Bud Black likely won on that basis. [http://hardballtalk.nbcsports.com/2012/11/13/how-to-win-the-manager-of-the-year-award-prove-the-pundits-wrong/]
Futuro verbo no gerúndio	simples +	# and secondly I know that Java compiler code optimization will take care of this anyway, i.e. the actual Java VM code will end up being inlined - so performance wise, there is no diffence between the two. [http://stackoverflow.com/questions/13380390/javascript-what-level-of-code-optimization-can-one-expect]
Presente verbo no gerúndio	simples +	This book contains two that rarely get attention, which is a shame. # I'm horrible at summing things up, but the books feature orphan/street-rate/teen Kim, originally hired to burgle a street magician but who ends up working with him instead to solve a magical mystery.
Passado verbo no gerúndio	simples +	# Can't speak to the food. I know that the long islands got me trashed. But that's their job. # I ended up buying 2 beers, 2 drinks, and a round (5) shots. Cost me \$65 before tip. [http://www.yelp.com/biz/the-living-room-boston]

Fonte: Dados da pesquisa

As buscas na seção resultaram em 11953 linhas de concordância para “*end up*” na seção *Web*. Foram selecionadas as linhas 1, 100, 200, 500, 1000 e 11952 para a realização do estudo.

De acordo com as amostras estudadas, “*end up*” é mais comum quando é seguido por um verbo no gerúndio, como em “*ended up calling*”, “*end up being*”, “*ends up working*” e “*ended up buying*”. Mais uma vez, os contextos são bem variados e com assuntos polêmicos que podem servir para reflexão e encorajar diálogo em língua-alvo na sala de aula, por exemplo, o segundo exemplo aborda a eutanásia, o assunto gera debates apimentados principalmente entre diversas religiões. No terceiro exemplo discutem-se as leis e a impunidade. O quarto exemplo traz questionamentos sobre como ocorre a premiação do gerente do mês. O quinto é mais técnico, pois trata do funcionamento do Java, um *software* de computador. O sexto traz uma resenha de um livro, e finalmente, o sétimo contexto trata de gastos com bebidas em Long Sland. Nesse sentido, novamente percebe-se que realmente é um PV com uma única acepção, mas com inúmeros campos de atuação na comunicação.

A fim de demonstrar mais uma potencialidade do COCA, efetuamos uma busca a partir de “END UP *ING” como resultado obtivemos uma lista com os 100 emparelhamentos mais frequentes de “*end up*”. O gráfico a seguir, ilustra os quatro emparelhamentos dessa natureza mais frequentes.

Figura 24 – Emparelhamentos mais frequentes de “*end up*” + verbo+ing (END UP *ING)

HELP		ALL FORMS (SAMPLE): 100 200 500	FREQ	TOTAL 34,644 UNIQUE 3,837 +
1	<input type="checkbox"/>	END UP BEING	2412	
2	<input type="checkbox"/>	ENDED UP BEING	1854	
3	<input type="checkbox"/>	ENDS UP BEING	983	
4	<input type="checkbox"/>	ENDED UP GETTING	681	

Fonte: COCA (2020)

Os resultados desta última busca evidenciam que “*end up*” se associa a outros verbos no gerúndio com frequência. Ao solicitar a busca com todas as letras maiúsculas, conseguimos a listagem lematizada, isto é, todas as palavras da família de “*end up*” serão agrupadas, o asterisco (*) seguido de “*ing*” apresentará qualquer verbo do gerúndio.

A lista completa totaliza 34644 ocorrências desta estrutura léxico-gramatical, sendo que a mais recorrente é “*end up being*” com 2412 ocorrências no corpúsculo. Esta análise mostra o quanto o COCA é uma amostra gigantesca e com inúmeras ferramentas de busca capazes de gerar dados para descrições linguísticas robustas, capazes de indicar tendências de usos de qualquer palavra do corpúsculo, inclusive dos PV, os quais são polissêmicos e idiomáticos, além

de produzirem diversos padrões de associação. Os PV são comuns na língua inglesa, portanto dominá-los requer o desenvolvimento da competência comunicativa intercultural. Nessa perspectiva encerramos nossa pesquisa tratando dos PV e da interculturalidade.

4.4A interculturalidade e os usos dos PV.

De acordo com Melo (2016, p. 6), “a língua é construída pelo uso que cada indivíduo faz dela, e esse uso é carregado de subjetividade, valores e crenças.” Nesse sentido, depreende-se que o conhecimento dos elementos culturais e sociais presentes no uso de uma língua é fundamental para se formar um falante intercultural capaz de transitar eficientemente entre sua língua-cultura e a língua-cultura alvo. Nesse sentido, com relação aos PV e os aprendizes brasileiros da língua inglesa, há uma relação de verdadeira aversão, pois não há estruturas lexicais semelhantes na língua inglesa, logo é mais fácil desprezá-los e aprender os verbos que são menos desafiadores e possuem carga de aprendizagem mais simples, pois os PV, em grande parte, são marcados pela imprevisibilidade semântica.

Para Kramsh (2012, p. 81), “no ensino de língua estrangeira, uma abordagem intercultural busca modos de entender o Outro do outro lado da fronteira através da aprendizagem de sua língua nacional.¹⁰⁹” Nesse sentido, é preciso compreender que estamos lidando com culturas diferentes, produzidas por comunidades diferentes, falantes de uma língua diferente e que lançam mão de palavras, estruturas léxico-gramaticais, governadas por regras sintáticas diferentes para se comunicarem.

Os gêneros têm suas próprias regras que influenciam nas escolhas lexicais e nos emparelhamentos a fim de executarem determinados papéis na sociedade, por exemplo, *point out* tende a ocorrer nas seções destinadas à explicação dos fundamentos teóricos em textos acadêmicos, ao passo que “*carry out*” também ocorre em textos da academia, mas nas seções que descrevem os procedimentos metodológicos. De acordo com a cultura norte-americana, as palavras formadas por uma única unidade lexical tendem a ser percebidas como mais formais que os PV, entretanto, esse fator não impede que PV desempenhem funções em textos da academia, tomados como mais formais e dotados de rigidez formal.

A linguagem pode se organizar de várias formas, porém, se um determinado falante faz uso de uma palavra ou estrutura de modo incomum, um falante nativo pode receber a

¹⁰⁹In foreign language teaching, a cross-cultural approach seeks ways to understand the Other on the other side of the border by learning his/her national language.

mensagem com certo estranhamento, isso se dá, porque embora certos emparelhamentos sejam possíveis, eles não ocorrem, realçando o aspecto probabilístico da linguagem.

Nesse sentido, conhecer os PV e seus contextos de uso é essencial para o desenvolvimento da competência intercultural, uma vez que dominar uma língua requer conhecimento das formas, de seus sentidos e, principalmente, de suas tendências de uso, isto é, ser consciente das restrições de uso latentes na linguagem. Não percebemos, mas como falantes da língua materna assimilamos naturalmente as regras de restrições de usos e as utilizamos incoscientemente.

No caso da língua-alvo, precisamos estudar estas regras, uma vez que não a utilizamos no cotidiano e em contextos naturais de fala. Uma das limitações do ensino de idiomas está relacionada à supervalorização do que é descrito como gramaticalmente aceito em detrimento do ensino e da aprendizagem dos padrões que realmente ocorrem. Por isso, é comum que aprendizes de língua inglesa se sintam frustrados ao saírem do centro de idiomas e, em sua primeira experiência de imersão na língua, sintam-se desconfortáveis, pois o “inglês dos livros” é muito diferente do “*real English*”.

Nessa perspectiva, preencher os livros didáticos destinados ao ensino da língua inglesa com textos autênticos, validados pelas pesquisas oriundas de *corpora*, pode contribuir para a superação das questões relacionadas à disparidade presente no inglês encontrado nos livros didáticos e o inglês falado em contextos naturais por seus falantes nativos.

Para exemplificar a pujança das pesquisas baseadas em *corpora*, propomos a comparação das informações demonstradas nas figuras 24 (seção anterior) e 25.

Figura 25 – “*End up*” + to be

HELP		ALL FORMS (SAMPLE): 100 200 500	FREQ	
1	<input type="checkbox"/>	ENDED UP TO BE	21	
2	<input type="checkbox"/>	END UP TO BE	15	
3	<input type="checkbox"/>	ENDS UP TO BE	9	
4	<input type="checkbox"/>	END UP TO BEING	1	
		TOTAL	46	

Fonte: COCA (2020)

Ao somarmos as ocorrências lematizadas de “*end up*” + verbo to be no gerúndio, temos como resultado 5249 ocorrências, mas quando somamos as ocorrências lematizadas do PV associadas com “to be”, encontramos apenas 46 ocorrências em todas as seções do *Corpus of Contemporary American English*. Por meio desta comparação de dados, pretendemos demonstrar que, embora possíveis, algumas associações são muito raras, enquanto outras são extremamente freqüentes.

Estes números demonstram que basear a produção de material didático em dados de pesquisas da LC pode contribuir substancialmente para a elaboração de livros que realmente retratam o uso da língua, promovendo contato do aprendiz com estruturas realmente relevantes para a comunicação. Logo, um falante intercultural, ao produzir as frases para se comunicar, não se baseia na sintaxe da língua materna, pois tem consciência de que sua língua-cultura é diferente da língua-cultura alvo. Desse modo, encerramos nossa argumentação sobre a interculturalidade e o uso dos PV, evidenciando a relevância do desenvolvimento da competência comunicativa intercultural, que requer o conhecimento das singularidades do “eu” e do “outro”, transitando entre o aqui e o lá sem perder a sua essência, mas acima de tudo respeitando as diferenças culturais, ao passo que faz uso coerente da linguagem, entendendo e se fazendo entender.

5 CONSIDERAÇÕES FINAIS

Ao tratar dos PV, procuramos apresentar uma visão quantitativa relacionada ao seu uso e distribuição na variedade da língua inglesa norte-americana, presente no COCA.

As pesquisas com *corpora* refletem e desvelam fenômenos linguísticos baseados no uso real da língua em estudo. Nesse sentido, a LC e o COCA proporcionam dados que podem refletir o comportamento autêntico do léxico, produzindo evidências.

Quanto aos resultados encontrados, no que se refere aos PV e seus comportamentos nos diferentes registros das amostras do *cópus*, em 2020, percebe-se que pode se encontrar um número maior de *types* em textos literários, assim lidar com textos dessa natureza em contextos de ensino-aprendizagem de idiomas pode oportunizar o contato com 53 dos 150 dos PV analisados. Caso, a escolha privilegie aqueles que se destacam em registros da TV e filmes, os aprendizes podem ter contato com 40 PV diferentes. Como 21 se destacaram em contextos da fala do cotidiano, é possível que ao levar para o âmbito escolar, réplicas de diálogos reais, o aprendiz possa assimilar formas, sentidos e contextos de usos, além de relacioná-las a estruturas léxico-gramaticais que constituem os diálogos.

Os textos da *internet* podem fornecer subsídios para 13 diferentes *types*, apesar de o número ser mais reduzido, pode ser importante privilegiá-los, pois a linguagem virtual vem se tornando cada vez mais presente no cotidiano de todos, e desempenha papéis importantes para a realização de diferentes práticas sociais em ambientes *online*. Em um mundo, onde é preciso estar sempre atualizado e informado, dominar vocabulário presente em jornais e revistas pode contribuir para a compreensão plena das mensagens, nesse sentido, 20 dos 150 PV de alta frequência se destacam em textos informativos.

No que se refere ao contexto da academia, o domínio de três PV podem auxiliar na divulgação de pesquisas acadêmicas. Portanto, percebe-se que conhecer os PV significa dominar aspectos relacionados não apenas às formas, seus sentidos e seus traços léxico-gramaticais, como também, em quais registros desempenham função relevante na linguagem.

A comparação entre o comportamento dos PV em 2011 e 2020 revelou que ao longo do período de dez anos ocorreram inúmeras mudanças nas tendências de uso, esse fato evidencia que a língua sofre pressões não apenas internas, mas também é influenciada por forças externas ao sistema.

No que se refere à primeira questão da pesquisa, “Quais são as expectativas dos documentos oficiais (BRASIL, 2002, 2020) relacionadas à formação de professores quanto ao conhecimento dos PV?”, considerando o ensino dos PV e as relações com as competências

ênfatisadas, os resultados obtidos nesta pesquisa, podem contribuir no sentido de que o conhecimento do comportamento dos PV em diferentes registros orais e escritos é um aporte importante para que o docente se conscientize de aspectos importantes das lexias em questão, ampliando o domínio do conteúdo, proporcionando condições para o desenvolvimento de práticas pedagógicas mais profícuas e gerando motivação e uma instrução baseada em fatos. A inclusão de informações baseadas em corpus no âmbito do ensino de idiomas preenche uma lacuna, ou seja, complementa o material com informações que podem tornar o ensino de línguas muito mais significativo.

O presente trabalho está em consonância com fundamentos balizadores da educação no âmbito universitário, contribuindo para o desenvolvimento de competências e habilidades endossadas pelos dois documentos analisados, ao aproximar teoria e prática, munindo os sujeitos do processo não apenas com conteúdo linguístico, como também, tornando-os aptos a empregar o conhecimento linguístico em suas práticas sócio-profissionais. Além do mais, conhecer o comportamento dos PV em diversos registros da linguagem pode contribuir para que aprendizes aproximem seu uso da língua-alvo ao desempenho real dos nativos.

Com relação à segunda pergunta, “Qual é a relação entre a Interculturalidade e o uso de PV?”, a comparação entre as frequências de uso dos PV nas variedades britânica e norte-americana evidencia que as tendências de uso podem ser influenciadas por fatores internos e externos da língua, uma vez que os falantes da língua utilizam os PV para desempenharem inúmeras funções comunicativas seja em contextos marcados pela formalidade, seja por contextos mais informais. Questões inerentes à subjetividade dos falantes e à funcionalidade da língua precisam ser consideradas no ensino e aprendizagem da língua-alvo, pois ao longo de dez anos entre a pesquisa de Liu (2011) e este trabalho, percebe-se que vários PV mantêm sua tendência, outros apresentaram aumento, enquanto alguns tendem a diminuir gradativamente seus usos. As pesquisas *cross-register* permitem apontar os contextos em que atuam, por exemplo, *point out* e *carry out* destacam-se em contextos mais formais da escrita acadêmica, porém *point out* se destaca nas seções destinadas à fundamentação teórica, enquanto *carry out* se destaca na descrição dos procedimentos metodológicos.

Tais disparidades nas tendências de uso demonstram que os aprendizes precisam adquirir consciência das relações culturais que permeiam a linguagem, de modo que aprendam a lidar com diversas culturas sem pressupor a superioridade de uma sobre a outra, e sim, tomando cada uma como algo singular, ao passo em que desperta a consciência do eu e do outro e a necessidade de se transitar lá e cá, através de uma postura de respeito às diferenças, a fim de

promover diálogos marcados pelo desenvolvimento da competência comunicativa intercultural.

A competência comunicativa intercultural demanda uma concepção de ensino de línguas votada para o domínio da língua, este requer não apenas a aprendizagem de formas e significados, mas acima de tudo é preciso compreender seus usos. Nesse sentido, os dados produzidos neste trabalho podem contribuir para um ensino intercultural de PV da língua inglesa.

Considerando a terceira pergunta, “Qual registro configura-se como um lócus de ocorrência de PV?” A seção *Fiction* se destaca como o principal lócus de ocorrência de PV, considerando as frequências por milhão mais elevadas de cada PV. O segundo lócus é seção *TV/Movies* com 40 PV, e o terceiro lócus é a seção *Spoken* com 21 PV se destacando. A seção *Fiction* é um importante contexto para se estudar PV na modalidade escrita, a seção *TV/Movies* é uma seção mista, pois os roteiros de filmes embora escritos, têm finalidade de serem oralizados na dramaturgia, enquanto a seção *Spoken* é um importante lócus para estudo de PV em contextos marcados pela oralidade, espontaneidade e informalidade.

No que tange à quarta pergunta “Como é o tratamento dos PV em dicionários, considerando os padrões de uso evidenciados pelo COCA?”. À medida que as investigações foram ganhando corpo, percebeu-se que os dicionários LDOCE e *Collins COBUILD*, ambos desenvolvidos a partir de pesquisas de *corpora*, trazem em algumas entradas marcações acerca de tendências de usos em contextos de oralidade, escrita e informalidade e formalidade. Os dados levantados na presente pesquisa demonstraram que a LC e seus preceitos metodológicos podem enriquecer os dicionários, por meio de acréscimos de novas acepções, e demarcações de usos em contextos distintos. Estas informações são fundamentais para que o dicionário se torne uma ferramenta para o desenvolvimento da competência comunicativa intercultural dos falantes de inglês em contextos de ESP, ESL e EFL.

Há aspectos na linguagem que precisam ser considerados, a saber, frequência, padrões de ocorrência e abrangência, ao realizar escolhas conscientes e eficazes em relação ao ensino de vocabulário da língua inglesa.

REFERÊNCIAS

- ALUÍSIO, S. M. ; ALMEIDA, G. M. de B. O que é e como se contrói um *corpus*? Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, São Leopoldo, v. 4. n. 3, p. 155-177, set./dez., 2006.
- BEDNAREK, M. **Language and television series: a linguistic approach to TV dialogue**. Cambridge: Cambridge University Press, 2018.
- BERBER SARDINHA, T. Linguística de Córpus: histórico e problemática. **Delta**, São Paulo, v. 16, n. 2, p. 323-367, 2000.
- BERBER SARDINHA, T. **Linguística de Córpus**. Barueri: Manole, 2004.
- BERBER SARDINHA, T. **A língua portuguesa no computador**. Campinas: Mercado de Letras, 2005.
- BERBER SARDINHA, T. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.
- BERBER SARDINHA, T. Como usar a Linguística de Córpus no ensino de língua estrangeira: por uma Linguística de Córpus educacional brasileira. *In*: VIANA, V. ; TAGNÍN, S. E. O. (org.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010. p. 293-348.
- BERBER SARDINHA, T. Linguística de Córpus. *In*: GONÇALVES, A. V. ; GÓIS, M. L. S. (org.). **Ciências da linguagem: o fazer científico?** v. 1, Campinas: Mercado de Letras, 2012. p. 321-347.
- BERBER SARDINHA, T. ; DELFINO, M. C. ; RAMPASO, M. Preparação de material didático para ensino de línguas com baseado em *corpora*. **The specialist**, São Paulo, v. 38, n. 1, p. 1-14, jan./jul. 2017.
- BIBER, D. ; CONRAD, S. ; REPPEN, R. Goals and methods of the corpus-based approach. *In*: BIBER, D. ; CONRAD, S. ; REPPEN, R. **Corpus Linguistics: investigating language, structure and use**. Cambridge: Cambridge University Press, 2004. p. 1-18.
- BIBER, D. ; RAPPEN, R. What does frequency have to do with grammar teaching. **Studies in second language acquisition**, Cambridge, v. 24, n. 2, p. 199-208, 2002.
- CANCLINI, N. G. **Culturas híbridas: estratégias para se entrar e sair da modernidade**. 4 ed. São Paulo: Edusp, 2015.
- CARTER, R. ; MCCARTHY, M. **Cambridge grammar of English: a comprehensive guide to spoken and written English usage**. Version 1.1. Cambridge: Cambridge University Press, 2006.
- CHAMBERS, A. Towards the corpus revolution? Bridging the research-practice gap. **Language teaching**, Cambridge, v. 52, 2019, p. 460-475.

DAVIES, M. The corpus of contemporary American English as the first reliable monitor corpus. **Literary and linguistic computing**, Oxford, v. 25, n. 4, p. 447-464, 2010.

DAVIES, M. *Corpus of Contemporary American English*. 2020. Disponível em: <https://www.english-corpora.org/coca>. Acesso em: 07 out. 2020.

DAVIES, M.; GARDNER, D. **A frequency dictionary of American English**: word sketches, collocations and thematic lists. Routledge: New York, 2010.

DEMO, P. **Educar pela pesquisa**. Campinas: Autores Associados, 2015.

FLOWERDEW, L. Applying Corpus Linguistics to pedagogy. **International journal of Corpus Linguistics**, Amsterdam, v.14, n. 3, p. 393-417, 2009.

FOX, C. *et al.* (ed.). **Longman dictionary of contemporary English**. 4th. ed. [S. l.]: Pearson Education Limited, 2003.

GARDNER, D. ; DAVIES, M. Pointing out frequent phrasal verbs: a corpus based analysis. **Tesol quarterly**, Hoboken, v. 41, n. 2, p. 339-359, 2007.

HARMER, J. **How to teach English**. 11th. ed. Essex: Pearson Educacional Limited, 2015.

KENNEDY, G. **An introduction to Corpus Linguistics**. [S.l.]: Pearson Education Limited, 1998.

LIU, D. The most frequently used English phrasal in American and British English: a multicorpus examination. **Tesol quarterly**, Hoboken, v. 45, n. 4, p. 661-688, 2011.

MACAULAY, A. ; SEATON, M. (ed.). **Collins Cobuild advanced learner's English dictionary**. 5th. ed. [S. l.]: HarperCollins Publishers, 2006.

MCCARTEN, J. **Teaching vocabulary**: lessons from the corpus, lessons for the classroom. Cambridge: Cambridge University Press, 2007.

MCENERY, T. ; BREZINA, V. ; GABLASOVA, D. ; BANERJEE, ; J. Corpus Linguistics, learner corpora, and SLA: employing technology to analyze language use. **Annual review of Applied Linguistics**, Cambridge, v. 39, n. 00, p. 74-92, 2019.

MCENERY, T. ; HARDIE, A. **Corpus Linguistics**: method, theory and practice. Cambridge: Cambridge University Press, 2012.

MCENERY, T.; WILSON, A. **Corpus Linguistics**: an introduction. 2nd. ed. Edinburgh: Edinburgh University Press, 2001.

MEUNIER, F. ; REPPEN, R. Corpus versus non-corpus-informed pedagogical materials: grammar as the focus. *In*: BIBER, D. ; REPPEN, R. **The Cambridge handbook of English Corpus Linguistics**. Cambridge: Cambridge University Press, 2015, p. 498-514.

MOON, R. Vocabulary connections: multi-word items in English. *In*: SCHMITT, N.; MCCARTHY, M. (ed.). **Vocabulary**: description, acquisition and pedagogy. Cambridge: Cambridge University Press, 1997. p. 40-63.

NATION, P. **Learning vocabulary in another language**. Cambridge: Cambridge university press, 2001.

NATION, P. **Como estruturar o aprendizado de vocabulário**. Tradução Cristiane Arruda. São Paulo: Special Books Services, 2003.

OLIVEIRA, L. P. Linguística de Córpus: teoria, interfaces e aplicações. **Matraga**, Rio de Janeiro, v. 16, n. 24, jan./jun, 2009, p. 48-76.

SARMENTO, S. Linguística de Córpus: histórico, metodologia, campos de aplicação. **Trama**, Marechal Cândido Rondon, v. 6, n. 12, p. 87-107, 2010.

SILVA, E. B. Jogos eletrônicos em língua inglesa: aspectos quantitativos do conteúdo lexical. **Via litterae**, Anápolis, v. 4, n. 1, jan./jun. 2012, p. 53-62.

SILVA, E. B. ; BABINI, M. A preparação de material terminológico em língua inglesa por meio de ferramentas linguístico-computacionais. **Trabalhos em Linguística Aplicada**, Campinas, v. 50, n. 01, p. 119-132, jan./jun. 2011.

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

SINCLAIR, J. **Trust the text**: language, corpus and discourse. London: Routledge, 2004.

VIANA, V. Linguística de Córpus: conceitos, técnicas & análises. *In*: VIANA, V. ; TAGNIN, S. E. O. (org.). **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial, 2010. p. 22-92.

APÊNDICE A – Lista de PV no COCA, frequência por milhão, em 2020

2020	2011	PV	ALL	BLOG	WEB	TV/M	SPOK	FIC	MAG	NEWS	ACAD
1	25	come on	224,94	31,48	31,49	1,409,49	124,46	110,60	21,10	23,51	4,83
2	1	go on	223,23	186,89	171,04	459,04	395,35	228,00	134,15	141,22	55,44
3	7	come out	112,62	109,64	98,90	135,73	244,70	134,14	72,93	80,33	19,63
4	3	come back	112,57	68,94	67,14	236,74	232,61	157,90	51,34	68,47	11,85
5	14	come in	109,28	73,08	70,55	214,12	174,30	133,74	85,53	92,84	25,29
6	13	get out	108,34	67,08	62,77	349,49	129,65	139,84	47,90	52,09	9,28
7	4	come up	108,34	97,84	83,67	127,72	280,52	111,70	68,09	72,23	18,91
8	5	go back	107,88	68,94	67,14	236,74	232,61	157,90	51,34	68,47	11,85
9	6	find out	94,37	90,04	94,05	194,42	148,44	87,83	65,82	46,28	20,85
10	2	pick up	87,49	69,82	66,90	114,37	82,55	192,01	80,23	77,24	19,11
11	8	go out	83,01	66,02	63,64	153,46	133,19	112,55	54,48	64,85	12,52
12	76	go through	77,20	85,44	74,60	99,99	154,91	62,57	51,72	63,86	19,64
13	18	end up	73,80	111,68	96,20	73,91	93,15	54,48	72,65	63,64	20,00
14	10	grow up	70,82	73,24	67,11	69,05	94,05	65,47	78,05	96,32	21,24
15	98	get in	67,17	52,73	44,67	183,32	92,62	68,74	39,55	42,49	7,66
16	12	turn out	65,99	74,69	71,82	63,08	82,82	66,73	80,39	56,55	29,70
17	9	point out	63,43	100,72	83,68	6,74	75,67	35,48	72,23	53,72	78,30
18	11	set up	57,71	64,67	65,22	46,82	53,88	44,89	73,97	72,44	38,74
19	19	get back	57,67	45,02	37,40	171,66	66,64	73,01	26,32	32,63	4,25
20	21	figure out	57,07	76,56	64,97	82,12	71,11	52,46	53,76	40,64	11,10
21	16	give up	56,71	66,57	59,53	79,57	54,49	62,65	46,40	62,40	20,34
22	17	make up	50,01	57,58	60,15	42,85	40,83	46,97	56,81	53,78	40,73
23	27	show up	49,66	55,20	45,58	80,49	56,91	51,71	46,79	49,33	8,67
24	54	go in	48,63	34,52	32,23	102,85	95,49	56,11	25,72	31,72	7,40
25	26	go down	47,60	41,55	38,26	91,54	78,26	64,06	28,14	29,97	6,66

26	22	sit down	47,28	22,84	28,41	99,42	57,30	115,32	27,52	21,82	6,30
27	35	wake up	46,68	47,12	49,62	109,44	38,59	70,64	29,34	22,10	4,14
28	23	get up	44,42	22,94	26,03	103,55	45,36	107,09	25,82	20,10	4,78
29	15	take on	41,19	48,21	44,56	17,15	53,42	27,81	53,86	49,71	34,20
30	33	go up	39,49	35,80	34,92	50,49	74,58	47,72	29,50	35,14	6,26
31	20	look up	38,77	22,04	28,43	21,71	17,81	188,68	19,55	14,29	4,83
32	29	work out	38,52	44,83	37,50	58,59	42,95	30,12	40,57	38,84	12,51
33	87	put on	37,75	33,07	32,31	42,44	47,77	64,60	36,61	34,93	10,37
34	30	stand up	37,10	33,02	34,61	37,26	51,14	87,38	25,09	22,45	7,02
35	31	come down	37,08	31,27	27,90	63,71	58,78	55,93	24,72	26,80	6,34
36	77	hold on	37,01	15,24	17,98	136,84	37,28	40,58	18,23	18,19	8,50
37	32	go ahead	36,78	16,99	14,28	106,85	107,15	24,17	9,27	8,69	2,29
38	37	take over	35,58	37,52	36,40	35,60	39,01	27,94	34,79	56,66	15,94
39	127	get on	33,72	28,65	23,89	75,42	51,95	39,04	21,19	23,42	3,97
40	50	move on	33,62	38,15	32,89	49,67	49,57	34,63	26,73	25,87	9,71
41	149	put in	31,60	39,47	34,45	29,97	47,95	28,65	27,92	30,26	12,86
42	49	check out	29,87	56,21	43,24	44,19	19,49	16,93	33,03	18,79	4,08
43	28	take off	29,57	19,88	18,04	47,79	30,13	67,65	26,04	22,61	5,03
44	56	get off	27,59	19,52	17,21	91,40	25,03	31,14	14,05	17,62	2,41
45	105	shut up	27,56	11,48	10,88	148,58	8,95	28,64	4,08	2,99	0,92
46	103	run out	26,61	24,85	22,60	40,79	32,33	39,92	21,94	24,93	4,96
47	34	look back	26,03	22,27	22,54	14,01	35,89	67,70	21,02	18,50	7,91
48	47	bring in	25,64	24,12	22,06	22,73	46,82	19,43	23,24	35,07	10,89
49	48	open up	25,23	25,62	23,04	39,34	31,43	19,85	25,79	19,44	16,09
50	90	move in	24,78	17,77	17,71	34,46	25,98	41,66	23,89	24,43	12,64
51	95	back up	24,65	26,92	24,27	35,39	24,74	42,96	20,10	17,68	4,84
52	36	carry out	24,41	20,60	32,81	6,08	22,37	8,48	20,38	23,85	62,08
53	53	catch up	24,40	26,31	23,31	32,54	24,33	35,16	23,17	21,66	8,31
54	46	look out	23,79	10,99	12,38	47,29	16,09	73,58	14,79	12,73	3,88
55	24	take out	22,81	22,61	21,58	28,22	26,88	44,84	16,95	16,65	4,92

56	39	pull out	22,11	15,02	14,75	15,73	23,92	67,41	18,24	19,66	4,08
57	41	take up	22,02	20,38	26,71	10,39	15,70	33,14	27,38	22,09	21,35
58	74	go over	22,00	15,67	15,70	46,32	34,18	37,69	11,10	11,33	3,30
59	38	hold up	21,32	13,28	14,16	32,63	17,93	53,58	17,54	16,08	6,28
60	86	hang out	21,07	20,06	17,40	55,56	18,22	20,25	17,52	15,03	2,79
61	60	go off	21,03	17,21	17,21	32,43	32,14	38,13	13,85	13,70	3,36
62	59	reach out	20,83	18,78	17,76	12,21	27,56	46,11	18,22	18,38	8,61
63	52	look round/around	20,55	10,91	11,48	24,16	12,84	86,17	10,43	8,53	2,64
64	89	come over	20,53	8,56	9,28	58,58	22,11	45,52	10,10	7,81	1,89
65	40	turn round/around	20,13	12,57	14,21	36,74	21,71	50,19	11,14	12,40	2,59
66	43	put up	20,04	27,21	19,90	22,76	24,62	22,15	16,46	22,15	4,29
67	134	take in	19,88	18,22	21,29	8,58	16,31	31,51	23,84	22,49	17,73
68	57	keep up	19,80	30,04	26,18	17,17	13,64	22,79	20,94	19,74	7,40
69	45	bring up	19,35	31,71	24,55	14,13	32,24	19,70	13,89	11,48	6,10
70	66	shut down	18,95	23,43	19,11	19,29	26,42	12,04	18,76	26,24	5,34
71	42	look down	18,90	7,93	11,82	11,63	9,31	93,55	11,14	6,26	3,14
72	80	break up	18,73	14,07	18,01	48,05	15,87	16,58	17,42	13,42	5,07
73	83	hang on	18,21	7,75	7,77	58,15	14,20	30,81	12,51	10,92	2,77
74	65	clean up	17,54	18,05	14,30	31,04	17,24	19,13	16,31	17,69	5,85
75	55	break down	17,42	19,03	19,18	12,76	18,54	12,53	24,73	16,53	15,69
76	117	get through	17,19	15,89	14,42	31,51	29,34	18,00	12,05	12,59	2,61
77	68	slow down	15,92	14,70	13,95	30,45	13,32	18,80	17,83	11,87	5,74
78	61	cut off	15,43	12,37	16,72	16,79	17,23	21,31	16,22	15,14	7,68
79	138	come through	15,36	9,91	10,33	29,67	20,31	24,42	10,72	12,96	4,29
80	51	put out	15,05	17,38	14,58	13,37	31,76	16,51	11,63	11,85	2,67
81	78	pay off	14,86	20,41	18,93	11,10	13,43	7,11	20,27	22,84	4,10
82	69	wind up	14,52	13,61	11,82	14,30	17,94	15,58	19,57	20,20	2,84
83	64	set out	14,43	10,80	16,85	4,40	11,23	18,92	23,41	14,07	16,38
84	71	line up	14,47	14,91	12,32	12,34	13,89	20,12	17,28	21,32	3,63
85	88	get down	14,20	6,38	6,09	53,83	14,98	16,53	6,87	6,18	1,35

86	115	go round /around	13,55	13,54	10,87	23,19	23,76	18,68	7,98	7,43	2,26
87	132	come off	13,50	17,66	14,08	12,82	13,61	13,27	12,65	22,43	1,08
88	73	lay out	13,48	13,65	14,06	5,01	22,10	15,64	16,11	12,69	8,56
89	106	turn off	13,28	13,18	14,62	13,95	8,31	28,67	15,21	9,36	3,30
90	112	come along	13,22	12,37	11,16	25,24	15,06	19,11	10,46	9,66	2,20
91	91	start out	13,01	16,86	15,45	9,34	19,57	11,06	15,29	12,23	3,71
92	92	call out	12,88	12,49	15,08	6,21	6,98	44,18	9,11	7,10	3,31
93	84	build up	12,73	15,87	15,98	6,28	14,86	8,10	19,84	10,83	9,70
94	70	turn up	12,50	8,53	9,33	14,45	11,24	22,73	16,65	12,97	4,33
95	102	keep on	12,36	14,80	14,68	18,16	13,28	18,12	9,44	7,46	2,55
96	75	hang up	12,30	5,12	6,49	22,49	6,98	44,35	7,72	5,35	0,93
97	99	blow up	12,12	11,80	11,40	23,25	17,17	11,55	9,18	9,61	2,18
98	104	make out	11,85	8,81	8,20	20,53	7,21	36,06	8,16	4,05	2,37
99	63	pull up	11,75	5,81	6,15	14,43	7,73	41,13	10,21	8,28	1,44
100	111	stand out	11,65	15,26	12,39	5,77	9,27	13,17	15,95	12,44	9,00
101	131	move out	11,61	8,20	8,09	24,77	11,19	17,41	8,11	11,23	3,64
102	44	bring back	11,46	13,54	12,31	12,25	14,58	12,61	10,31	11,60	4,14
103	113	play out	11,15	15,13	11,30	4,51	19,89	5,70	11,40	13,46	7,43
104	114	break out	11,15	9,42	11,07	10,21	12,42	13,46	12,95	13,49	6,18
105	144	go along	11,13	13,61	11,63	9,73	21,29	11,87	7,67	9,35	3,47
106	129	hand over	11,11	6,65	8,72	8,12	8,22	37,79	8,38	8,91	3,33
107	108	step back	11,04	9,03	8,21	10,17	11,00	34,17	7,60	6,15	2,88
108	62	turn back	10,85	4,48	6,70	6,58	5,41	51,92	6,35	4,85	2,57
109	133	pass on	10,84	13,69	12,35	8,09	9,23	11,54	12,53	11,76	7,47
110	143	give in	10,60	10,27	16,32	6,10	7,12	13,49	9,48	6,23	16,30
111	100	carry on	9,90	10,22	12,26	10,81	8,79	13,74	9,16	7,68	6,58
112	85	throw out	9,65	13,51	9,94	8,85	14,44	8,24	7,05	12,31	2,41
113	79	hold out	9,56	5,22	6,12	6,79	5,50	37,30	6,63	6,79	3,50
114	93	sit up	9,51	2,29	5,13	7,08	4,64	49,45	5,46	2,82	1,11
115	81	bring out	9,50	8,37	8,60	8,65	13,30	13,78	10,26	7,94	5,12

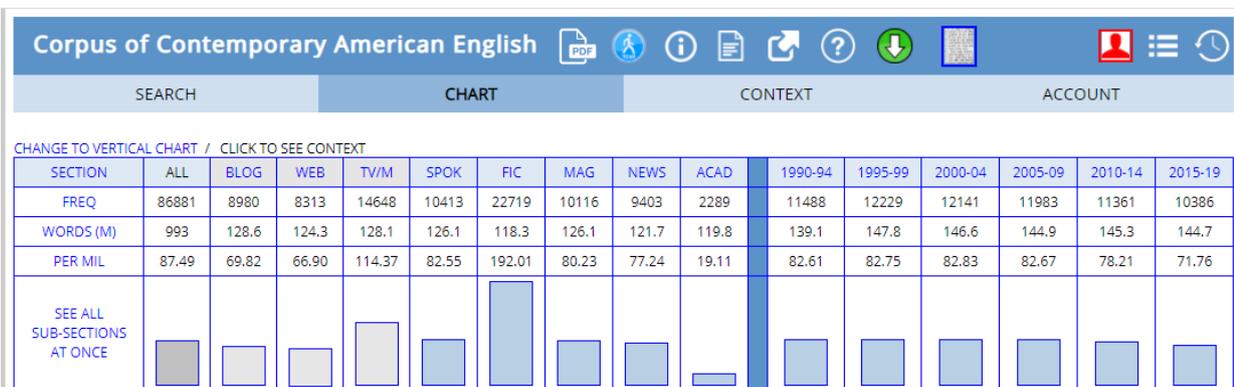
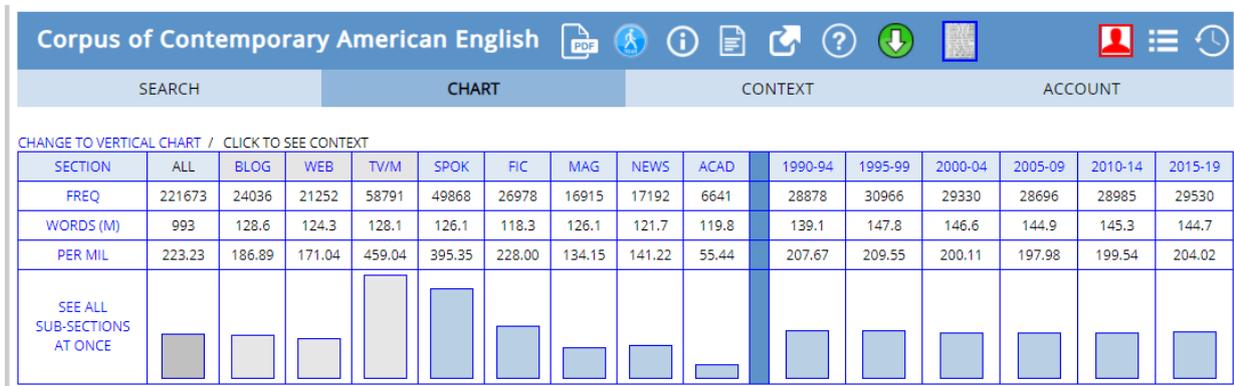
116	141	fill in	9,39	10,96	11,14	7,23	7,96	9,16	12,52	8,87	7,14
117	109	lay down	9,09	6,65	10,39	11,53	7,84	18,42	8,62	4,95	4,52
118	101	set off	8,98	5,29	7,23	5,44	7,59	16,83	13,49	12,51	3,92
119	116	walk on	8,92	6,78	7,36	11,28	6,94	19,72	8,60	8,07	2,94
120	107	bring about	8,90	10,56	12,22	1,63	7,94	2,54	7,95	5,77	22,92
121	137	follow up	8,54	13,61	11,57	4,54	13,24	2,48	7,68	6,06	8,68
122	58	put down	8,49	6,90	7,77	11,29	6,76	20,71	6,80	5,88	2,13
123	121	fill out	8,44	10,33	10,25	8,58	6,56	8,51	7,88	9,89	5,37
124	140	come around/round	8,40	6,61	6,24	16,53	8,82	17,38	5,07	5,31	1,18
125	67	turn over	8,34	4,68	6,34	5,74	12,68	12,49	8,19	12,81	4,04
126	94	turn down	7,94	6,55	6,85	7,16	7,68	11,96	9,30	12,07	2,09
127	122	sit back	7,78	7,56	6,19	8,07	6,82	23,23	5,00	5,10	0,78
128	124	move up	7,51	6,58	6,07	6,99	7,31	9,96	8,30	11,83	3,11
129	120	move back	7,50	6,24	6,03	10,96	6,37	13,07	6,57	8,49	2,35
130	82	pull back	7,22	3,20	3,90	4,39	7,88	27,04	5,88	5,21	1,24
131	123	rule out	6,92	6,08	7,20	5,02	8,58	2,16	8,18	9,57	8,53
132	126	take down	6,70	8,93	7,63	9,14	7,05	8,30	5,81	5,14	1,34
133	125	pick out	6,64	6,17	5,42	8,61	5,74	12,95	7,54	4,76	2,02
134	139	settle down	6,52	4,29	4,43	14,31	3,62	14,44	5,68	3,87	1,55
135	119	write down	6,28	7,21	8,72	7,13	4,68	8,97	6,02	3,18	4,29
136	118	hold back	6,27	6,35	6,36	5,46	6,64	11,95	5,28	5,82	2,46
137	110	bring down	6,24	6,97	7,77	5,22	9,39	5,59	5,79	6,36	2,65
138	130	sum up	6,19	8,93	9,09	1,13	4,11	2,49	9,13	6,56	7,98
139	142	give out	5,47	7,10	6,16	5,78	6,52	7,00	4,31	5,25	1,53
140	72	take back	5,19	6,40	6,04	7,03	6,83	4,98	4,11	4,24	1,60
141	147	come about	4,89	6,20	6,54	2,15	7,89	2,91	4,62	3,98	4,72
142	136	sort out	4,42	5,19	4,43	3,28	4,97	4,44	4,89	4,43	3,70
143	128	give back	4,15	5,98	5,20	3,44	5,29	2,41	3,95	5,27	1,48
144	146	put off	4,02	5,13	5,28	1,91	3,59	4,89	4,72	5,27	1,39
145	145	break off	3,87	1,90	3,36	3,17	2,60	11,20	4,43	3,06	1,63

146	148	close down	3,27	3,16	3,10	3,19	5,87	2,78	2,91	3,22	1,81
147	135	set down	2,67	0,96	2,12	1,09	0,98	12,84	1,80	0,96	1,19
148	150	set about	1,87	1,24	2,02	0,34	0,76	3,77	3,28	1,93	1,79
149	96	put back	1,68	2,06	1,81	1,37	2,50	2,04	1,39	1,56	0,67
150	97	send out	0,11	0,07	0,09	0,14	0,13	0,22	0,15	0,07	0,02

Fonte: Dados da presente pesquisa, com base no COCA.

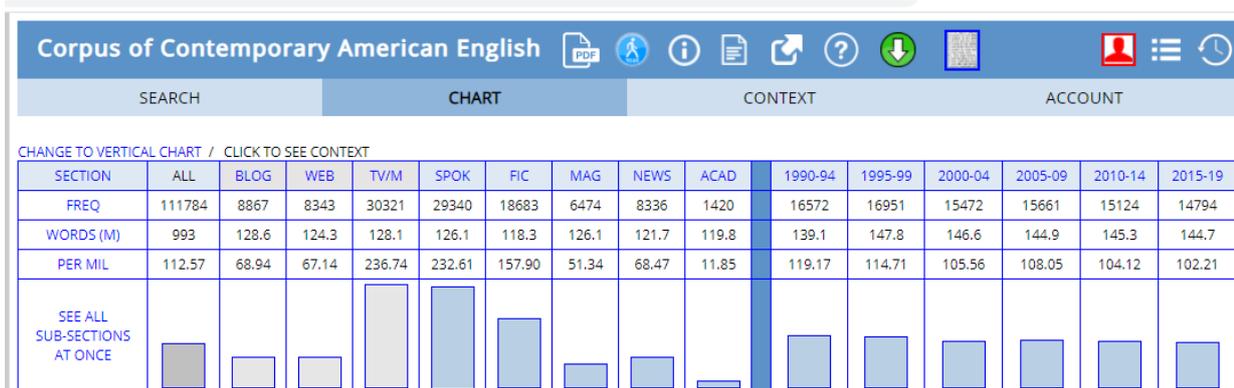
APÊNDICE B – Gráficos de frequência (COCA 2020)

GO ON

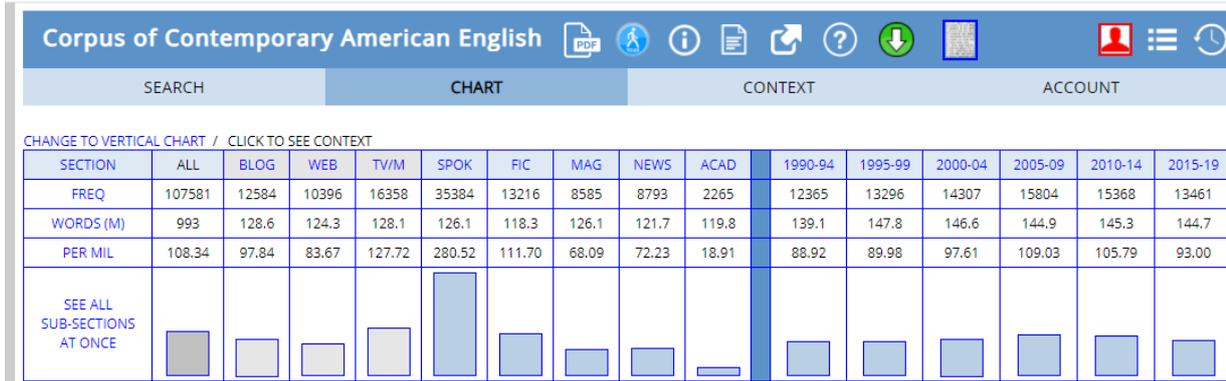


PICK UP

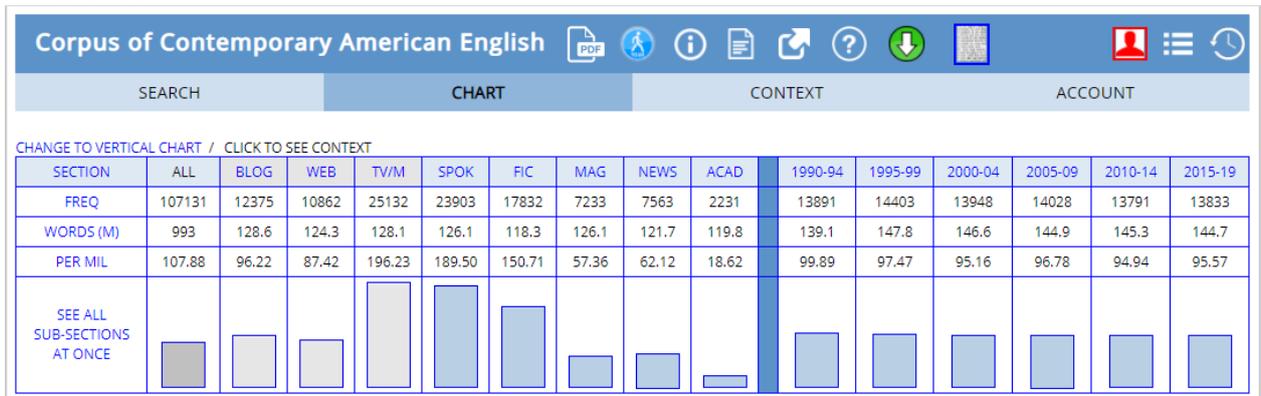
COME BACK



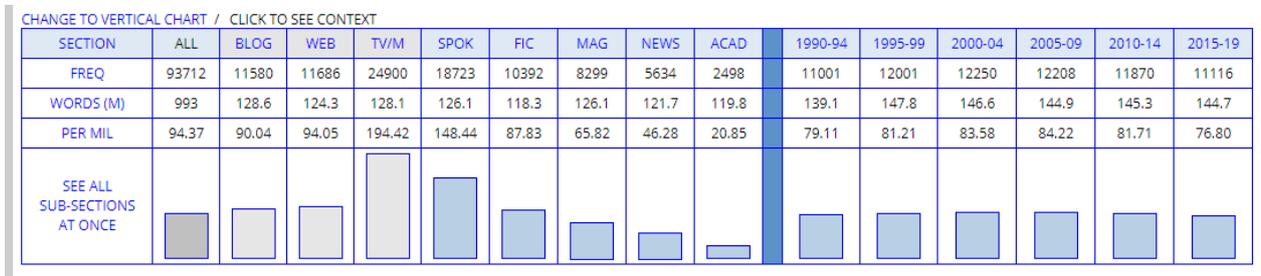
COME UP



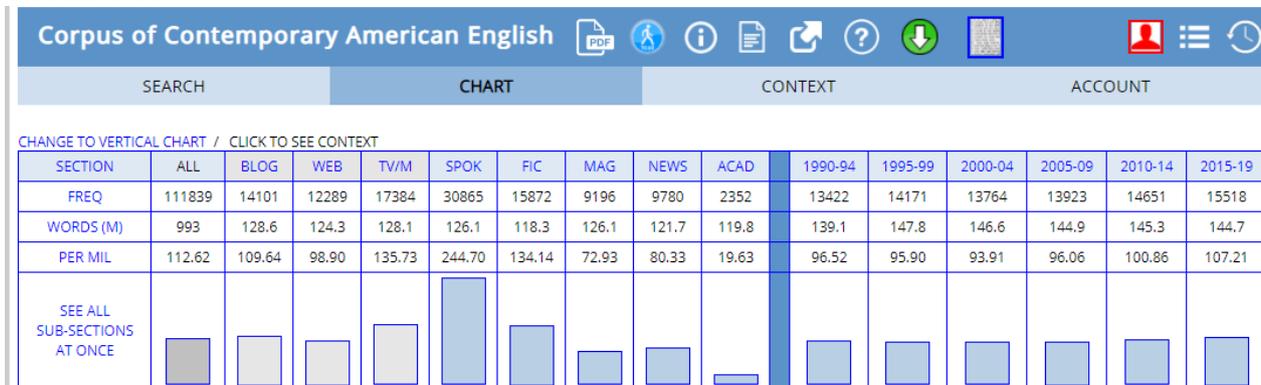
GO BACK



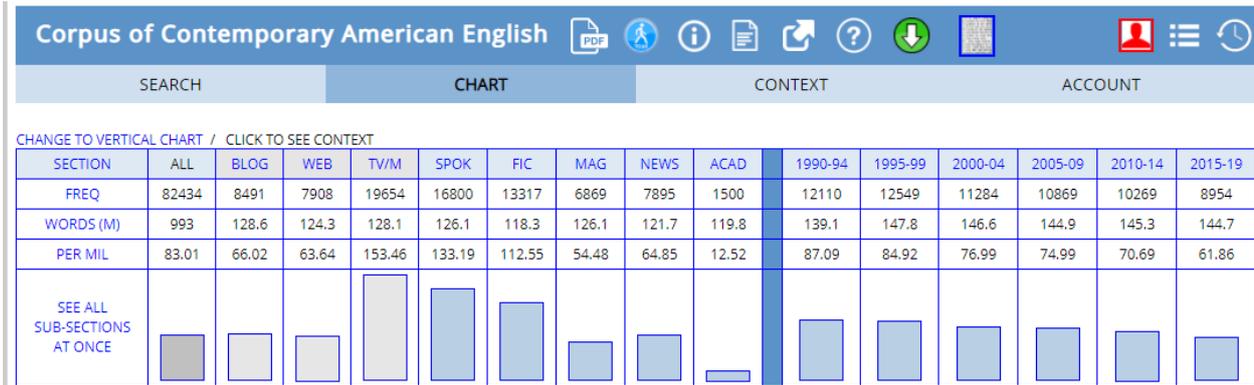
FIND OUT



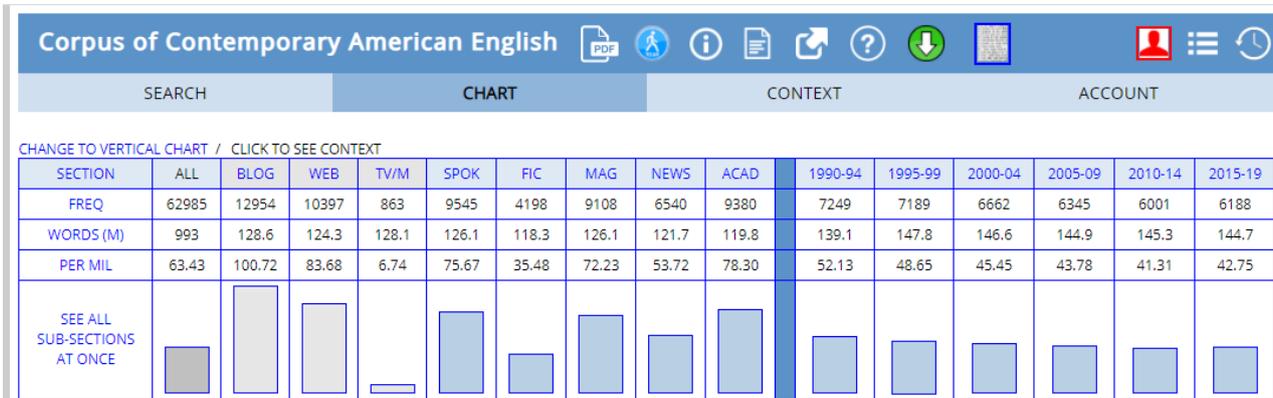
COME OUT



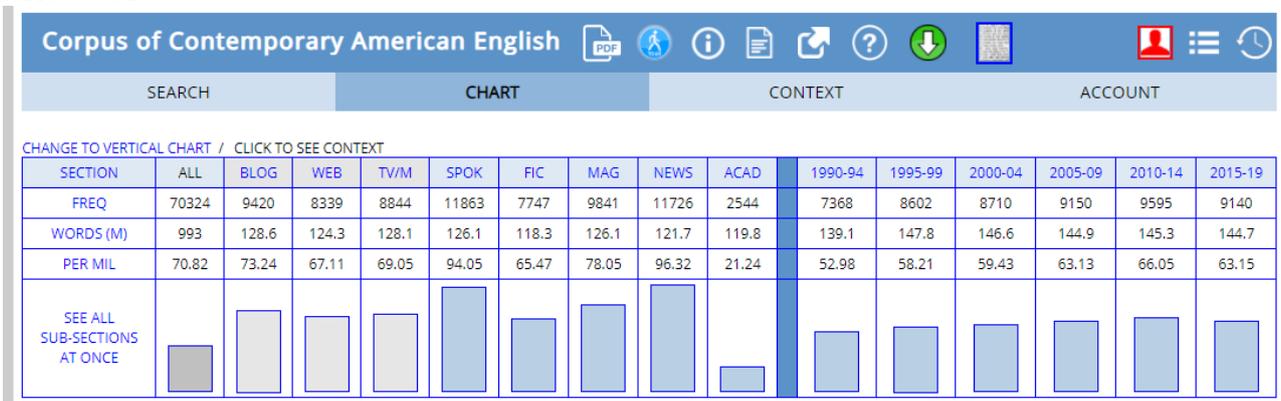
GO OUT



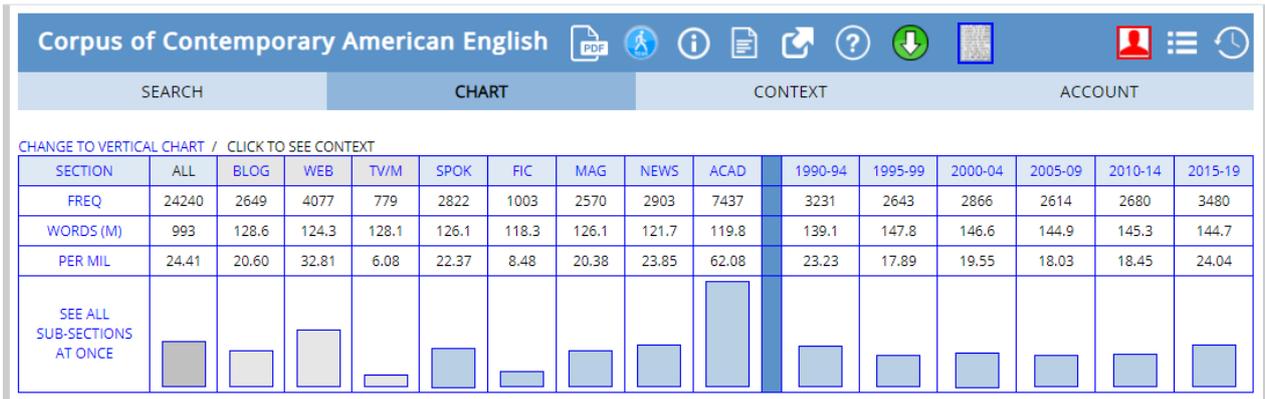
POINT OUT



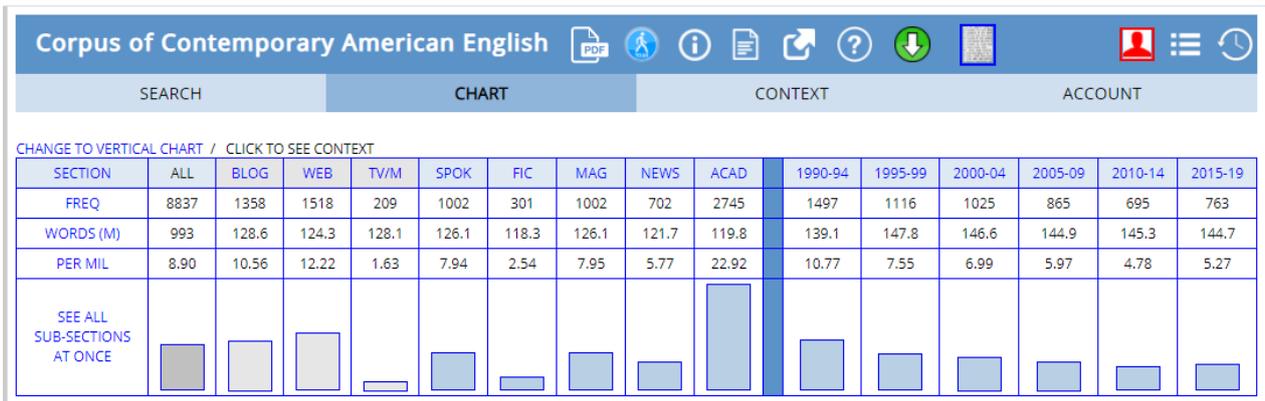
GROW UP



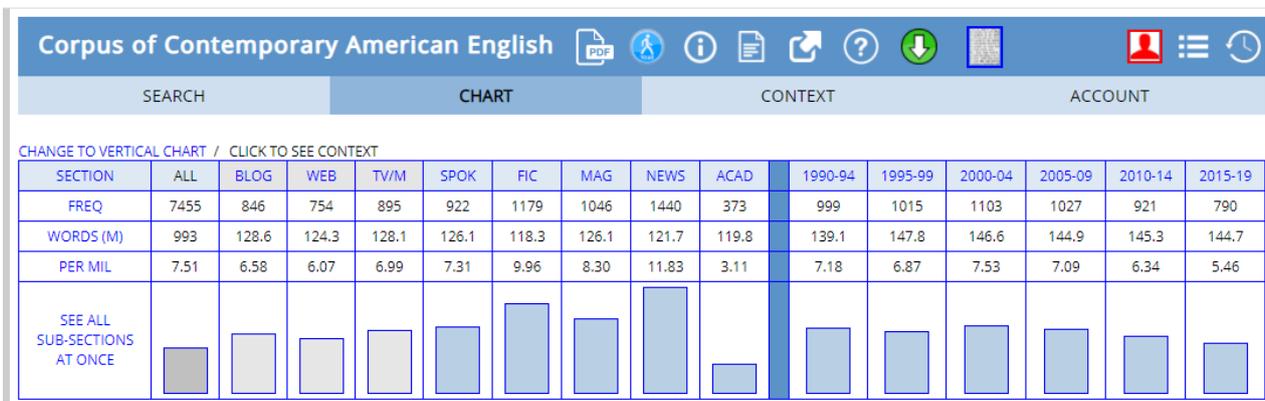
CARRY OUT



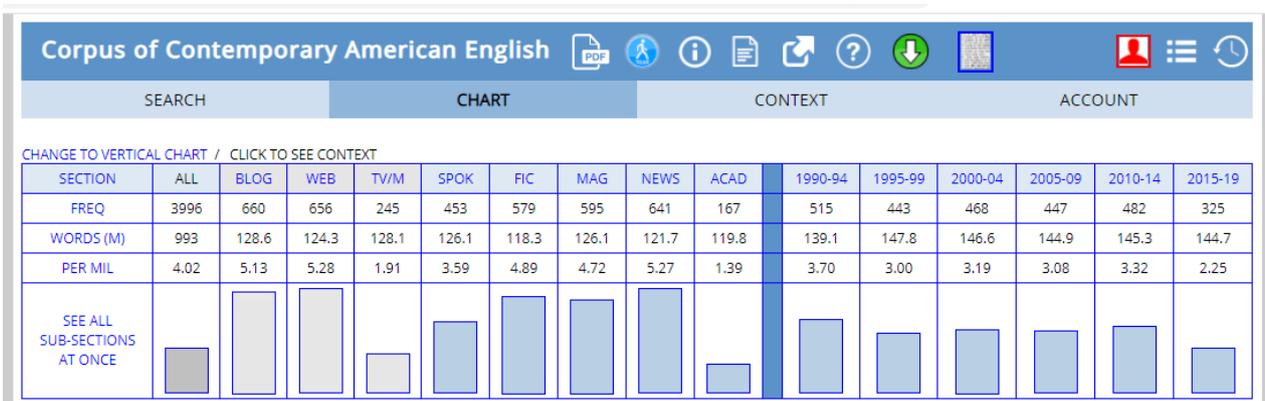
BRING ABOUT



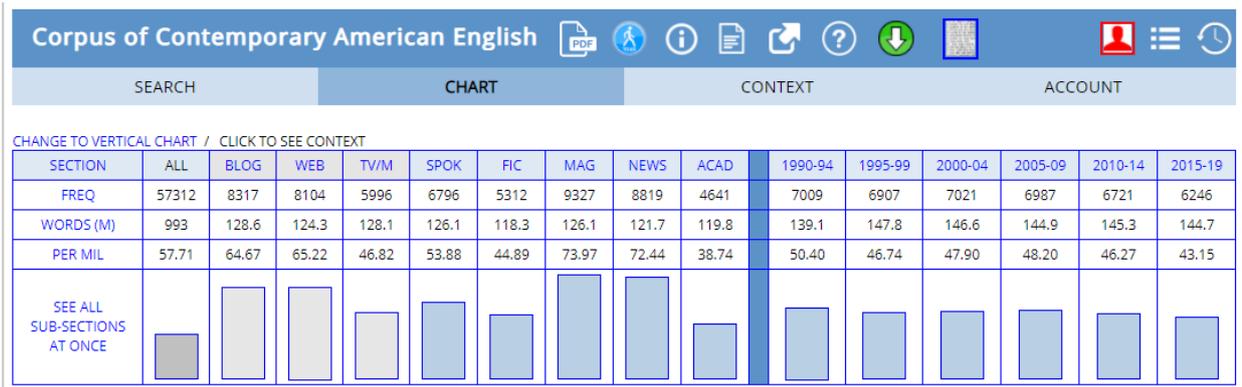
MOVE UP



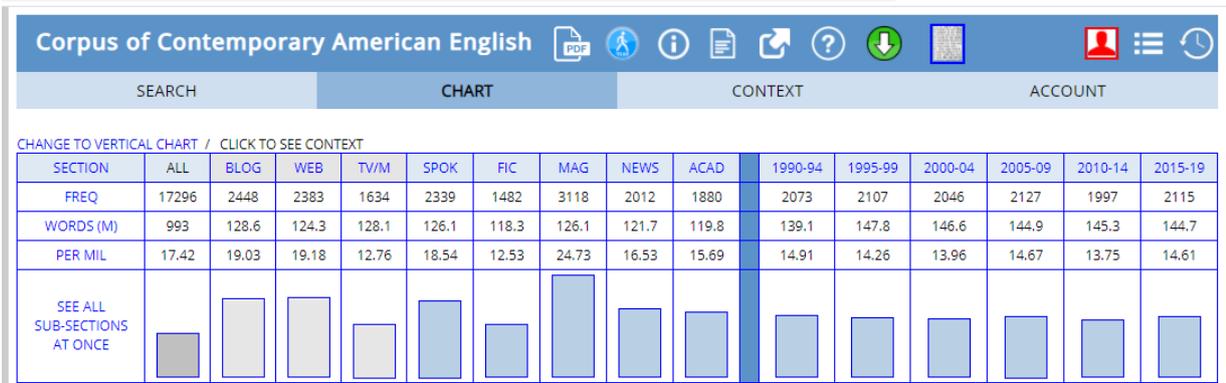
PUT OFF



SET UP

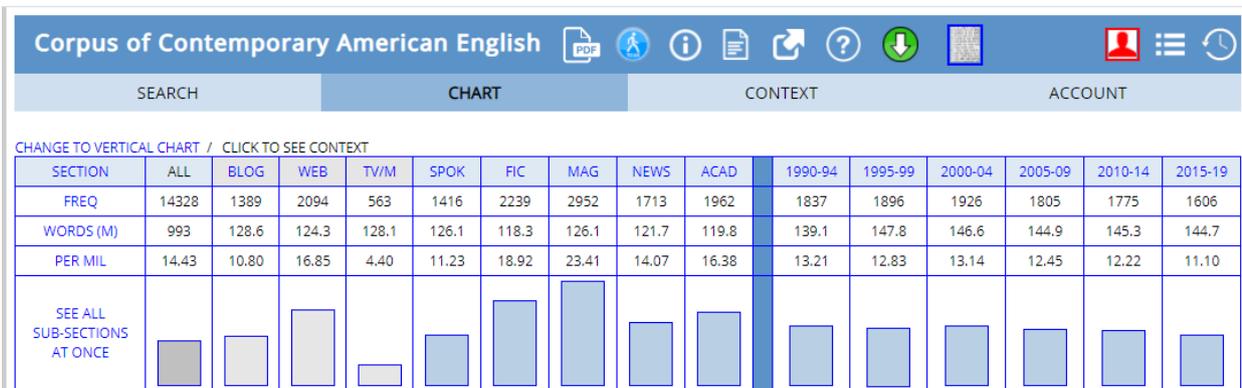


BREAK DOWN

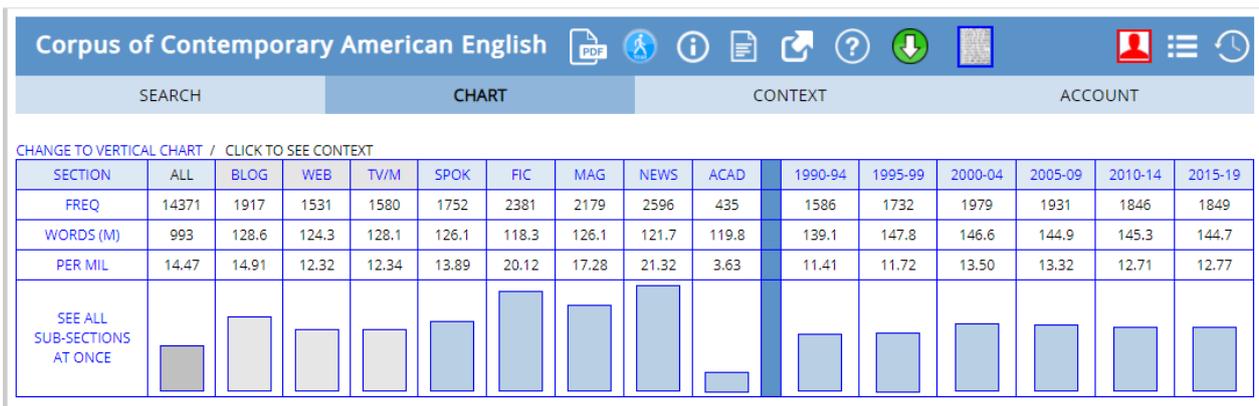


SET

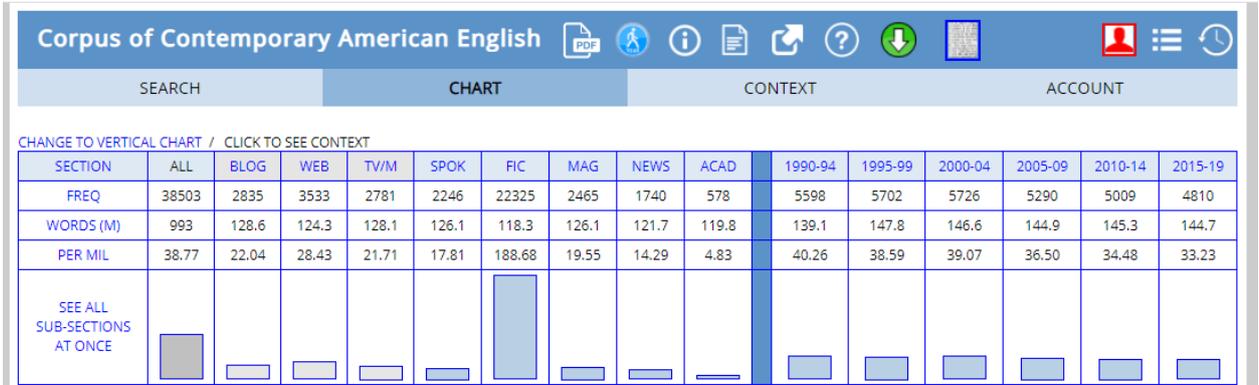
OUT



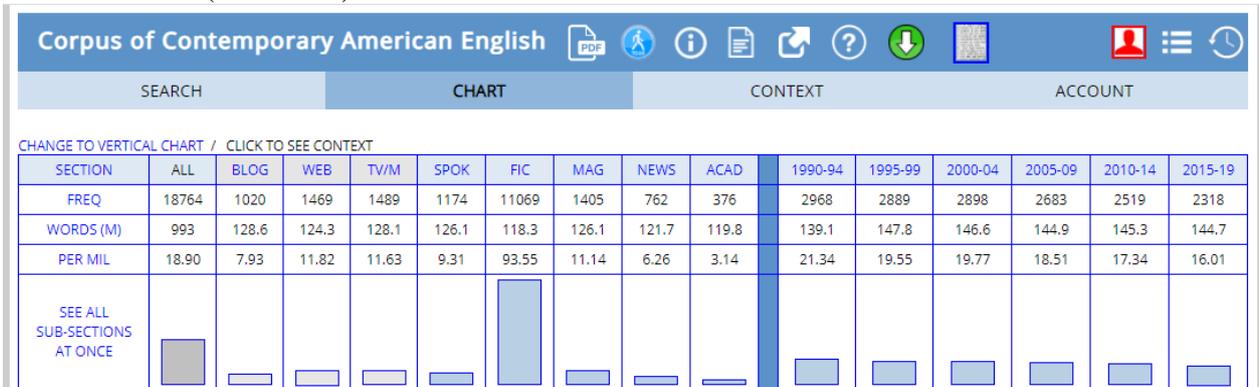
LINE UP



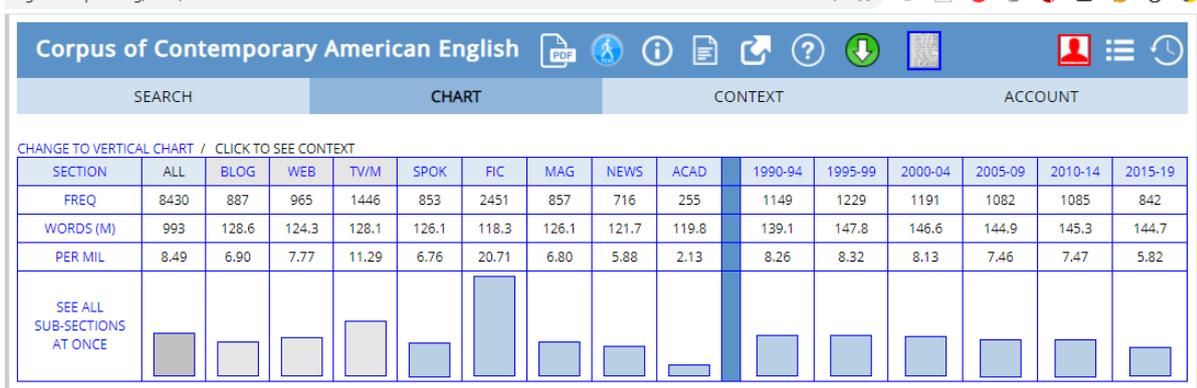
LOOK UP (FICTION)



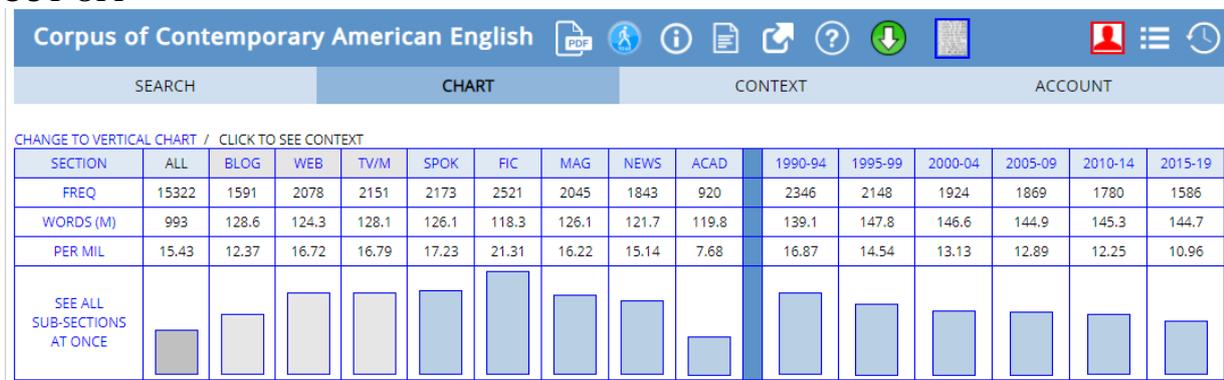
LOOK DOWN (FICTION)



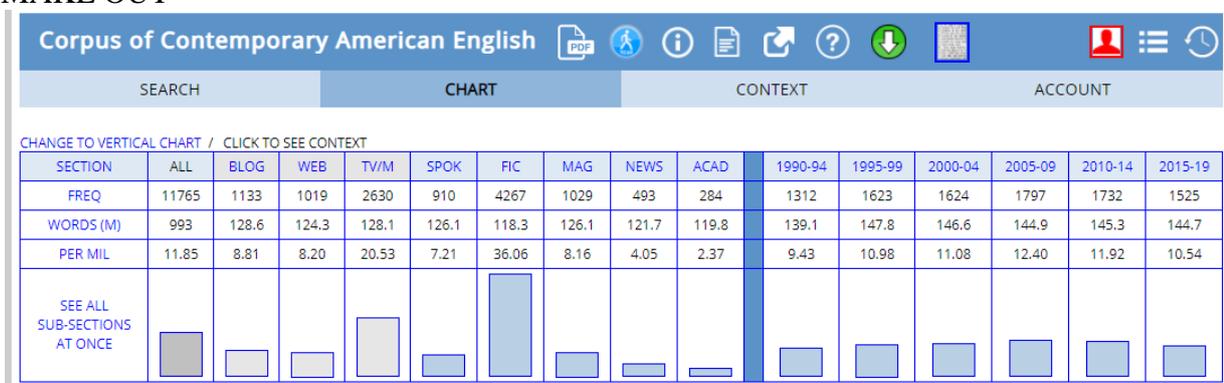
PUT DOWN (FICTION)



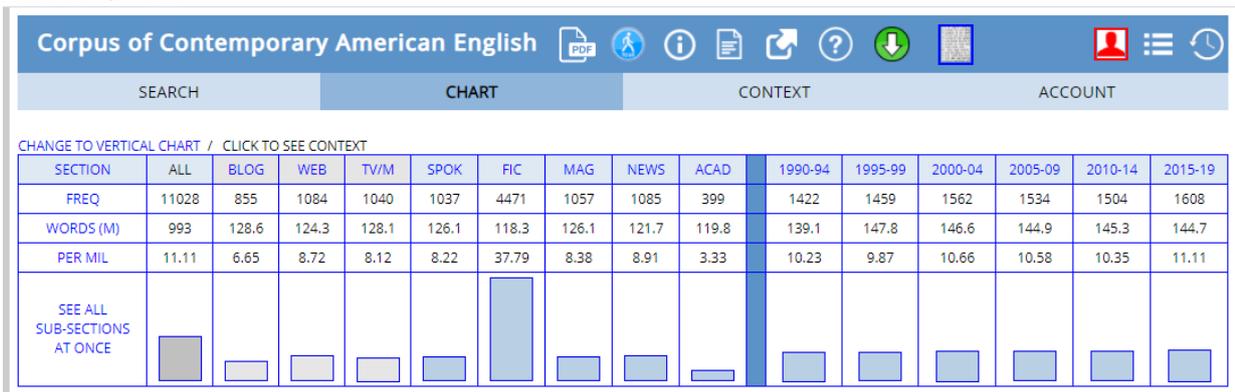
CUT OFF



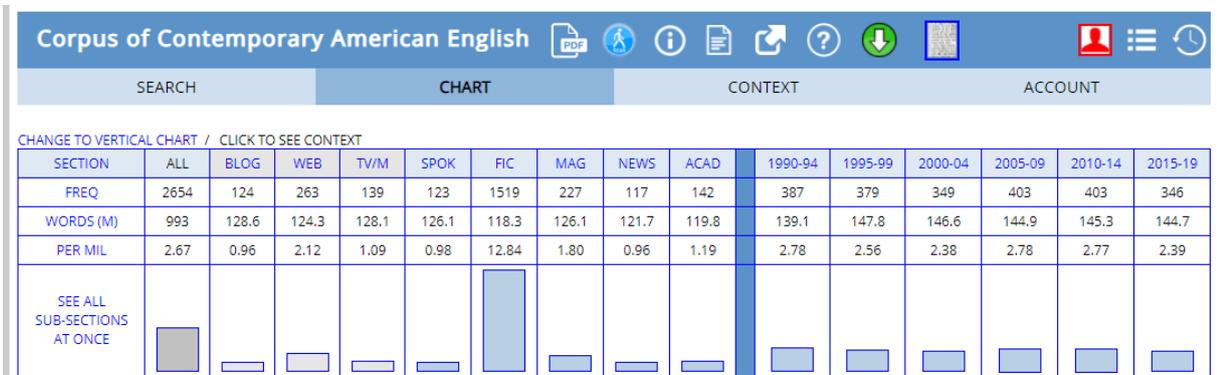
MAKE OUT



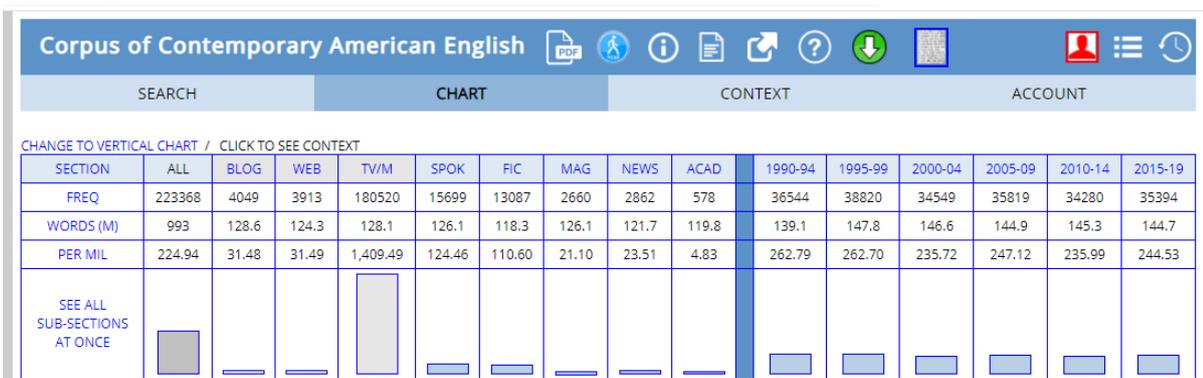
HAND OVER



SET DOWN

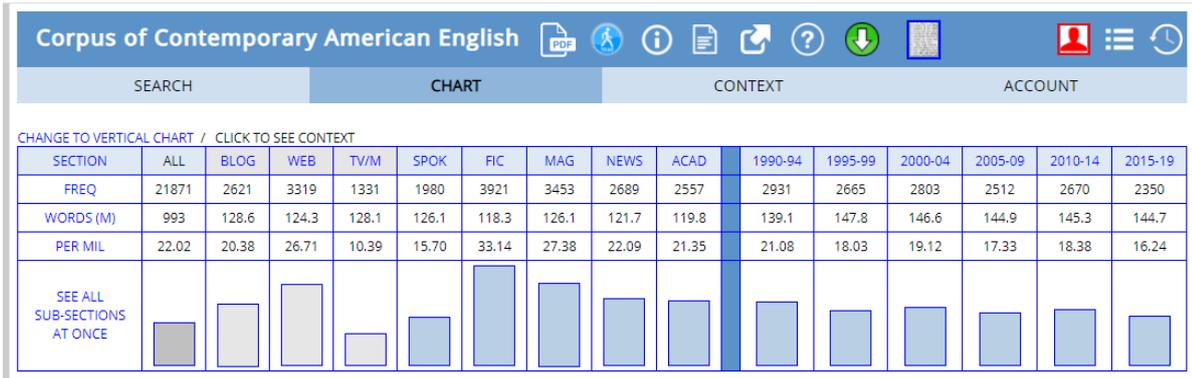


COME ON

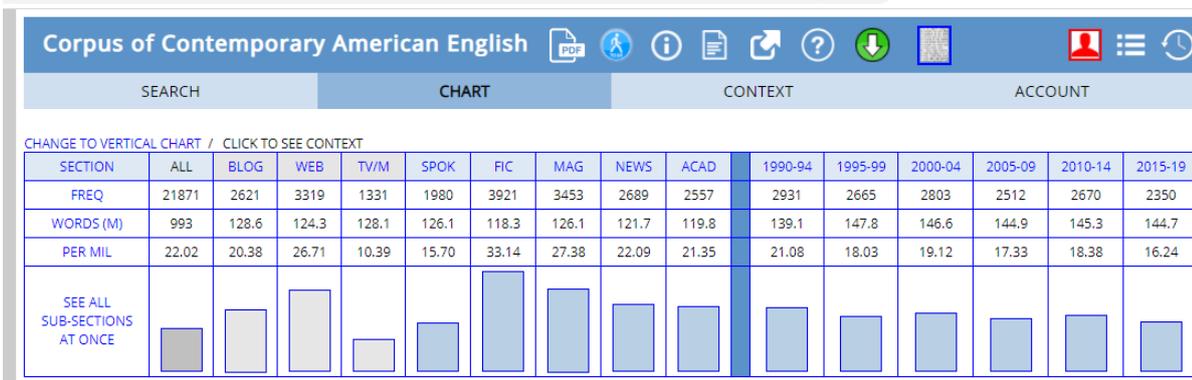


TAKE

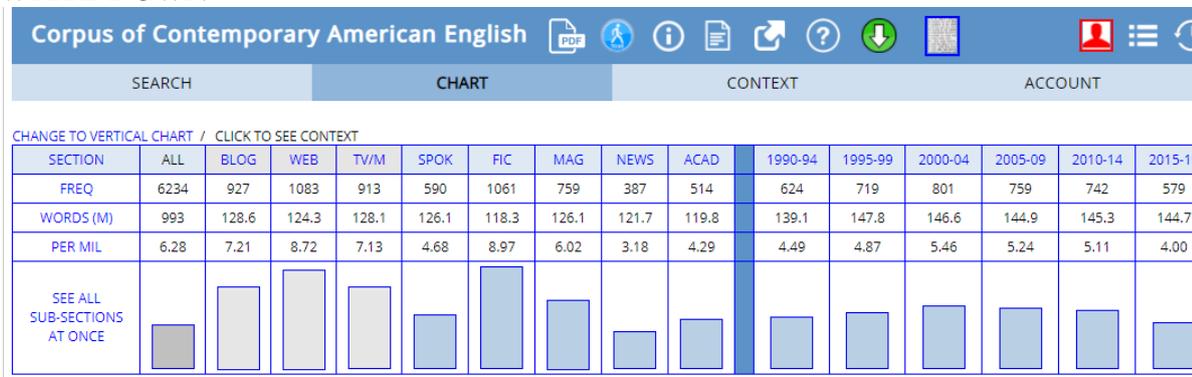
UP



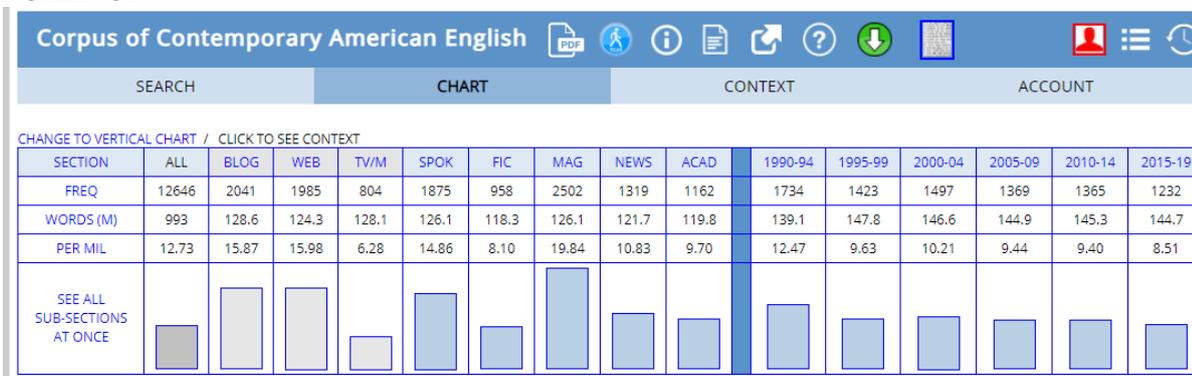
PAY OFF



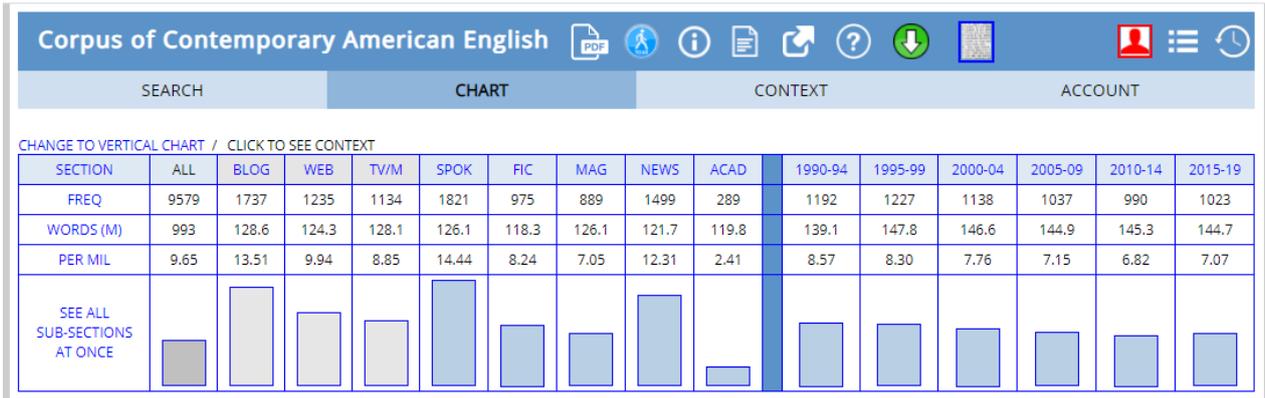
WRITE DOWN



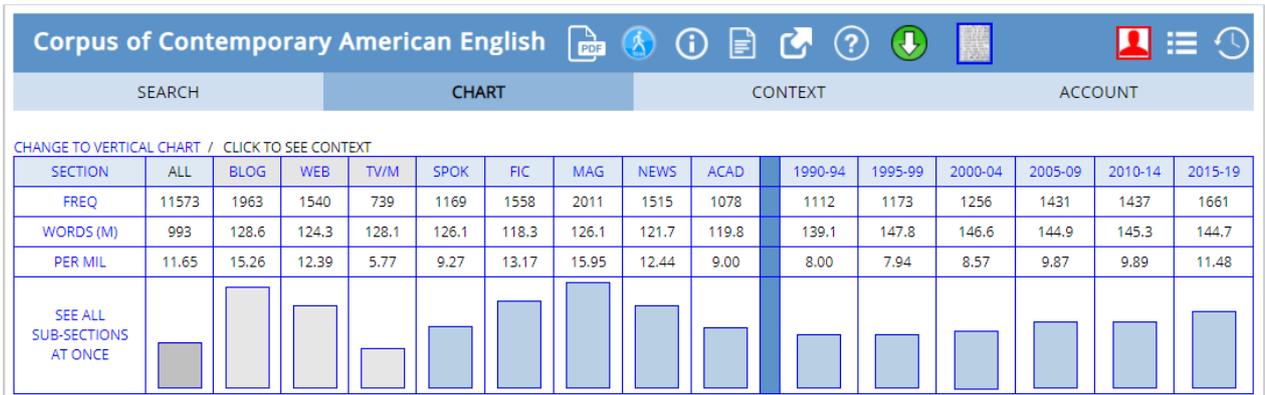
BUILD UP



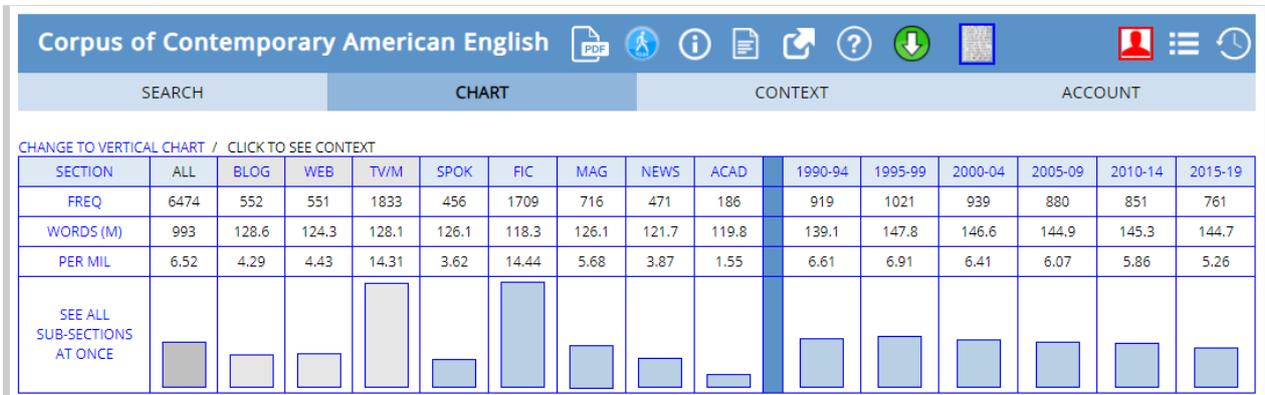
THROW OUT



STAND OUT



SETTLE DOWN



SET ABOUT

