

Multi-word verbs na escrita acadêmica de aprendizes do inglês: um estudo baseado em *corpus* de estudantes brasileiros

Priscilla Tulipa da Costa*

Adriana Maria Tenuta de Azevedo**

RESUMO

Motivada pela importância dos verbos compostos por mais de uma palavra na comunicação em inglês e pelos problemas que envolvem o ensino dessas formações lexicais, esta pesquisa teve como objetivo principal investigar o uso dos *multi-word verbs* em textos acadêmicos escritos por aprendizes brasileiros da língua inglesa. A proposta é identificar quais são os verbos frasais, preposicionados e frasais preposicionados mais comuns nos textos de aprendizes, bem como observar as diferenças e semelhanças no emprego dessas unidades em gênero textual argumentativo por parte de brasileiros e falantes nativos. Para tanto, a pesquisa encontrou suporte teórico na Linguística de *Corpus*, que “se ocupa da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais, em formato legível por computador, que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” (SARDINHA, 2004, p. 3). A investigação foi realizada com base em dois *corpora*: um de estudo, composto de textos escritos por aprendizes brasileiros do inglês (BRICLE) e um de referência, que contém textos de nativos da língua inglesa (LOCNESS). Os resultados sugerem que, na comparação com os nativos, os aprendizes brasileiros empregam quantidades bem significativas de *multi-word verbs* em seus textos. Entretanto, nota-se que existem erros na forma de uso, especialmente no que concerne às produções textuais dos não nativos.

Palavras-chave: Linguística de *corpus*. *Multi-word verbs*. *Phrasal verbs*. *Prepositional verbs*. Ensino e aprendizagem de idiomas.

1. Introdução

Os verbos compostos por mais de uma palavra (*multi-word verbs*) são construções gramaticais muito peculiares na língua inglesa. Amplamente utilizadas no dia a dia dos falantes nativos, essas estruturas lexicais tornaram-se

* Universidade Federal de Minas Gerais (UFMG).

** Universidade Federal de Minas Gerais (UFMG).

muito importantes para qualquer pessoa que deseje obter proficiência no idioma. Dominá-las, no entanto, não é tarefa fácil para a maioria dos estudantes, em especial para os aprendizes brasileiros, que encontram vários desafios na aprendizagem em razão de fatores sintáticos e semânticos (FARIA; BERNARDO; SILVA, 2007).

Tendo em vista que a fluência na língua inglesa oral e escrita torna-se imprescindível no mundo globalizado e que existem dificuldades no ensino/aprendizagem de estruturas essenciais da língua, como os verbos frasais e preposicionados, este trabalho busca mapear o uso dos *multi-word verbs* na linguagem escrita de gênero argumentativo. O objetivo é identificar como essas locuções são utilizadas por aprendizes brasileiros, se a forma de emprego pode influenciar na eficácia da comunicação e quais são as similaridades e diferenças percebidas na comparação com textos produzidos por nativos da língua inglesa.

Para tanto, optou-se por buscar na Linguística de *Corpus* o suporte necessário para a investigação. Foram usados dois *corpora*: um brasileiro, Br-ICLE, para estudo (CE), e um inglês, LOCNESS, para referência (CR), ambos contendo coleções de textos produzidos em língua inglesa. A análise dos dados foi realizada com o auxílio do *software Antconc*, que gerou as listas de concordância responsáveis por tornar a averiguação possível.

O trabalho é apresentado em três etapas principais: a primeira é constituída pela fundamentação teórica, na qual estão relacionadas as áreas de estudo que forneceram embasamento teórico para a pesquisa; a segunda, relata a metodologia utilizada na extração e análise dos dados; a terceira, contém os relatos da investigação, bem como os resultados obtidos por meio da pesquisa.

2. Fundamentação teórica

A pesquisa realizada tem fundamento em dois campos de estudo específicos que serão abordados nesta seção: os verbos compostos por mais de uma palavra (*multi-word verbs*) e a Linguística de *Corpus*.

2.1 Verbos constituídos por mais de uma palavra (*multi-word verbs*)

Os verbos constituídos por mais de uma palavra são construções muito comuns na língua inglesa, principalmente em discursos informais. Na literatura, eles são

denominados mais frequentemente como *multi-word verbs* e muitos autores, como Carter e McCarthy (2006), os dividem em três categorias principais: verbos frasais (*phrasal verbs*), verbos preposicionados (*prepositional verbs*) e verbos frasais preposicionados (*phrasal-prepositional verbs*). A seguir, apresentaremos uma breve definição de cada uma dessas subcategorias.

2.1.1 Multi-word verbs

De acordo com Carter e McCarthy (2006), os *multi-word verbs* são combinações entre verbos lexicais e partículas – que podem ser um advérbio ou uma preposição – formando uma única unidade de sentido, como em *This calls for a celebration* (*calls for* = *demands/requires*) e *The fridge has fallen apart* (*fallen apart* = *broken into pieces*). Para os autores, as palavras que formam os *multi-word verbs* são escritas separadamente, nunca como uma única palavra ou com hífenes.

Para Greenbaum e Quirk, “as duas principais categorias de *multi-word verbs* [*phrasal verbs* e *prepositional verbs*] consistem de um verbo lexical acrescido de uma partícula, designação neutra para as categorias de advérbio e preposição que são usadas em tais combinações.” (GREENBAUM; QUIRK, 1990, p. 336, tradução nossa). Quirk *et al.* acrescentam que os *multi-word verbs* são formados por “partículas morfologicamente invariáveis que, juntamente com o verbo, funcionam como uma única unidade gramatical” (QUIRK *et al.*, 1985, p. 1150, tradução nossa), ou seja, comportam-se como um único verbo, seja lexicalmente ou sintaticamente.

Acerca deste aspecto, em que verbo e partícula têm valor de um único verbo, Downing e Locke afirmam que “a função de muitas das partículas é modificar a natureza da atividade expressa pelo verbo. O resultado é uma extensão de sentido que é frequentemente diferente do (s) sentido (s) dos verbos sozinhos.” (DOWNING; LOCKE, 2002, p. 336, tradução nossa). Já Biber *et al.* (1999) afirmam que essa combinação normalmente tem sentido idiomático, e seu significado não pode ser deduzido a partir do significado individual de cada palavra.

Na classificação de Biber *et al.* (1999), os *multi-word verbs* estão divididos em quatro categorias distintas: (1) *Phrasal verbs* (verbos frasais); (2) *Prepositional verbs* (verbos preposicionados); (3) *Phrasal-prepositional verbs* (verbos frasais preposicionados); (4) *Other multi-word verb constructions* (outras construções de verbos compostos por mais de uma palavra).

Palmer (1974) denomina os *multi-word verbs* como *compound verbs*, que são compostos por um verbo + combinação de partículas (estas podem ser preposições, advérbios ou advérbios preposicionados) e suas classificações são: (1) *Phrasal verbs* (verbos frasais); (2) *Prepositional verbs* (verbos preposicionados); (3) *Related constructions* (construções correlatas). Por “construções correlatas”, Palmer (1974) entende verbo + construções de partículas (ou *phrasal prepositional verbs*) e verbos + outros elementos, como adjetivos (ex.: *He **made clear** his intentions*), infinitivos (ex.: *He **let slip** the opportunity*) e preposição seguida de *noun phrase* (*It **brings to light** the facts*).

Para realizar este trabalho, adotaremos a perspectiva de Carter e McCarthy (2006), que também é a mesma de Greenbaum e Quirk (1990) e Quirk *et al.* (1985). Com base na proposta destes autores, conheceremos, a seguir, as definições de cada um dos três tipos de *multi-word verbs*.

2.1.2 Verbos frasais (*phrasal verbs*)

Na concepção de Carter e McCarthy (2006), os verbos frasais consistem de um verbo lexical e de uma partícula e podem ser transitivos ou intransitivos. Os *phrasal verbs* intransitivos não requerem objeto, como em *get back*, *wake up* e *hang on* (Ex. ***Hang on.** Let me write all this down / Hang on = wait*). Os transitivos, por sua vez, necessitam de objeto direto, que pode vir antes ou depois da partícula (desde que o objeto não seja um pronome pessoal). Como exemplo, temos *Someone **turned** the light on*, que também pode ser escrito como *Someone **turned on** the light*.

2.1.3 Verbos preposicionados (*prepositional verbs*)

Os verbos preposicionados são assim classificados por serem uma junção entre um verbo lexical e uma preposição. Uma das características dessa locução é a exigência de um objeto direto, como em ***Look at** these pictures* e *I don't **care for** Jane's parties*. Carter e McCarthy (2006) explicitam que os verbos preposicionados seguem regras diferentes das que regem os verbos frasais porque, no caso dos *prepositional verbs*, o objeto direto (ainda que seja um pronome) deve seguir a preposição, como mostrado no exemplo extraído de Carter e McCarthy (2006, p. 434):

*I couldn't sell my car. I just couldn't **do without** it.*
*(I couldn't sell my car. I just couldn't **do it without**.)*
*(couldn't **do without** = need/have to)*

Para Greenbaum e Quirk, os “verbos preposicionados são superficialmente semelhantes aos verbos frasais transitivos, no entanto, há diferenças tanto sintáticas como fonológicas.” (GREENBAUM; QUIRK, 1990, p. 340, tradução nossa).

2.1.4 Verbos frasais preposicionados (*Phrasal-prepositional verbs*)

Carter e McCarthy (2006) definem os verbos frasais preposicionados como verbos lexicais seguidos por uma partícula adverbial e uma preposição, como em *catch up on*, *get along with*, *look out for* e *put up with*. Esses verbos ocorrem normalmente em contextos informais, são inseparáveis em sua estrutura e, uma vez que terminam em preposição, exigem sempre objeto direto, como em:

*I'm **looking forward to** the weekend.*¹ (*looking forward to* = esperar com entusiasmo)

*He had to **put up with** a lot of teasing at school.*² (*put up with* = tolerar)

Por fim, pode-se concluir que os verbos formados por mais de uma palavra podem ser classificados, de maneira geral, conforme a tabela abaixo:

Tabela 1: Principais tipos de *multi-word verbs*

	Verbo Lexical	Objeto direto	Partículas		Objeto direto
			Advérbio	Preposição	
<i>Phrasal verbs transitivos</i>	<i>Turn</i>	<i>Someone</i>	<i>Down</i>	-	-
<i>Phrasal verbs intransitivos</i>	<i>Crop</i>	-	<i>Up</i>	-	-

1 Exemplo extraído de Carter e McCarthy (2006, p. 436).

2 Exemplo extraído de Quirk *et al.* (1985, p. 1160).

<i>Prepositional verbs</i>	<i>Come</i>	-	-	<i>Across</i>	+ <i>a problem</i>
<i>Phrasal-prepositional verbs</i>	<i>Put</i>	-	<i>Up</i>	<i>With</i>	+ <i>his bad temper</i>

Fonte: Tabela adaptada de Greenbaum e Quirk (1990, p. 1181).

2.2 Multi-word verbs e o ensino da língua inglesa

Apesar de serem marcas essenciais na linguagem cotidiana dos falantes de língua inglesa e de fundamental importância para a boa comunicação no idioma³, os *multi-word verbs* são considerados alguns dos elementos gramaticais mais complicados de serem aprendidos por estudantes de inglês como segunda língua (ESL/EFL). Celse-Murcia e Larsen Freeman (1999) afirmam que poucas línguas não germânicas apresentam formações como os *phrasal verbs* (*multi-word verbs* em nossa terminologia), por isso, a maioria dos alunos acha esses verbos estranhos e difíceis. Entre os desafios que permeiam a aprendizagem desses elementos estão questões sintáticas e semânticas, como a idiomaticidade, ou seja, o fato de o significado da locução não ser, na maioria das vezes, compreendido a partir do significado individual dos termos que a compõem.

Em relação a isso, pesquisadores como Dempsey, McCarthy e McNamara (2007) relatam que é frequente aprendizes driblarem a idiomaticidade usando verbos “latinizados”, palavras que, apesar de equivalentes, não têm a mesma conotação dos verbos frasais. Esse fator provavelmente se torna um problema já que, conforme afirmam Celse-Murcia e Larsen Freeman (1999), os verbos frasais são ubíquos no inglês e ninguém pode falar ou entender o idioma, pelo menos no contexto informal, sem conhecer esse tipo de registro.

Por outro lado, Waibel (2007) mostra que os verbos latinizados podem parecer

3 Sobre a potencialidade dos *multi-word-verbs* para a comunicação em língua inglesa de não nativos, Bywater (1969 *apud* CORNELL, 1985, p. 270, tradução nossa) afirma: “O fato é que o que distingue a escrita e, acima de tudo, a fala de um bom estudante estrangeiro das de um falante nativo do inglês é que o que o falante nativo escreve ou diz está cheio dessas expressões [*multi-word verbs*], ao passo que a maioria dos aprendizes não nativos tem medo dessas estruturas, cuidadosamente as evitam e, consequentemente, soam inaturais quando as usam. Estudantes não nativos que gostam de ser lisonjeados por seu inglês podem alcançar isso usando corretamente um grande número desses verbos compostos”.

mais eruditos para os aprendizes não nativos e, por soarem mais formais, são mais utilizados. Para a autora, os aprendizes tendem a acreditar que o uso de palavras mais “sofisticadas” faz com que eles se pareçam mais proficientes na comunicação em inglês, aproximando suas *performances* às dos nativos da língua. Entretanto, o resultado disso é que os alunos nem sempre têm consciência de que “o uso de verbos latinizados é inadequado em determinadas situações.” (WAIBEL, 2007, p. 37, tradução nossa).

Em suma, a importância dos *multi-word verbs* para a proficiência na língua inglesa e as questões que envolvem seu ensino/aprendizagem são fatores que motivam pesquisadores em todo o mundo a escolherem essas expressões como objeto de estudo, seja no âmbito pedagógico ou descritivo. A maioria dos trabalhos nessa área é internacional, como o de Waibel (2007), que realizou análises quantitativas e qualitativas sobre o uso de verbos frasais nos textos em língua inglesa escritos por estudantes alemães e italianos, e descreveu os atuais problemas que os aprendizes avançados têm em relação ao emprego desses verbos. Seus resultados contribuíram para a compreensão de aspectos gerais da linguagem de aprendizes avançados.

No Brasil, podemos citar como exemplo os trabalhos de pesquisadoras como Vicentini (2006) e Rampazzi (2008), que analisaram o uso de *chunks* em *corpora* de linguagem oral constituído por trechos de seriados de TV. Ambos os trabalhos tiveram como foco a análise de padrões léxico-gramaticais para contribuição na elaboração de atividades didáticas.

2.3 Linguística de Corpus – definição

Uma das áreas investigativas que tem demonstrado grande relevância para os estudos do léxico e da linguagem é a linguística de *corpus*, campo da Linguística que “se ocupa da coleta e exploração de *corpora*, ou conjunto de dados linguísticos textuais, em formato legível por computador, que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística” (SARDINHA, 2004, p. 3). Trata-se de um trabalho cujo quadro conceitual é formado por uma abordagem empirista e que entende a linguagem como um sistema probabilístico (SARDINHA, 2000).

Pelo termo ‘*corpus*’ entende-se, na definição de Ducrot e Todorov, como um “conjunto tão variado quanto possível de enunciados efetivamente emitidos por

usuários da referida língua em determinada época.” (DUCROT; TODOROV, 2001, p. 42). McEnery e Wilson (2001) afirmam que, em princípio, qualquer coleção de mais de um texto pode ser chamada de *corpus*, já que este termo é o correspondente latino para ‘corpo’. Assim, um *corpus* pode ser definido como qualquer corpo de texto. Quando usada no contexto da linguística moderna, no entanto, a palavra *corpus* apresenta conotações mais específicas, que transcendem essa simples definição e que implicam amostragem e representatividade, tamanho finito, formato eletrônico e referência padrão.

A característica de formato eletrônico citada por McEnery e Wilson (2001) também aparece na concepção de Baker:

Hoje, o *corpus* significa, principalmente, uma coletânea de textos agrupados em formato eletrônico e capazes de serem analisados automática ou semi-automaticamente de diversas formas; um *corpus* não está mais restrito a textos escritos, mas inclui também textos falados e ele pode conter um grande número de textos de diversas fontes, produzidos por vários escritores e falantes e acerca de diferentes assuntos. O importante é que sejam compilados com um propósito e de acordo com critérios específicos de maneira a assegurar a representatividade de determinada área ou a amostra da língua a qual se pretende explicar. (BAKER, 1995, p. 225, tradução nossa).

Dessa forma, em concordância com a perspectiva de Sinclair, pode-se concluir que “um *corpus* é uma coleção de trechos de texto, em formato eletrônico, selecionados conforme critérios externos para representar, tanto quanto possível, uma língua ou variedade da língua como fonte de dados para pesquisa linguística” (SINCLAIR, 2004, tradução nossa).

2.3.2 Breve histórico da Linguística de Corpus

Apesar de não ser uma forma de estudo recente, como ressalta Sardinha (2004) ao relatar que já na Antiguidade e Idade Média eram produzidos *corpora* da Bíblia, foi somente após o advento da tecnologia que a linguística de *corpus* se tornou uma abordagem mais precisa.

Durante o século XX, muitos trabalhos foram realizados com base em pesquisas de *corpora*, entretanto, os processos de coleta e análise de *corpus* eram realizados manualmente e demandavam uma equipe muito grande de analistas. Esse fator,

aliado ao gigantesco número de dados de alguns trabalhos, favorecia o aumento das possibilidades de erros e de falhas na definição de padrões classificatórios. Da mesma forma, os resultados tornavam-se pouco confiáveis. Sardinha (2000) cita que em 1897, Kädling utilizou um *corpus* de 11 milhões de palavras para analisar a distribuição ortográfica do alemão, e que para realizar a tarefa foram necessários cinco mil analistas. Em 1921, Edward Thorndike realizou um trabalho revolucionário para o ensino de língua inglesa, que abordava a frequência de palavras no idioma. A investigação, realizada manualmente, foi baseada em um *corpus* composto por 4,5 milhões de palavras, e anos mais tarde foi revisada com a aplicação de um *corpus* ainda maior, com 18 milhões de palavras.

A informatização das informações e o uso do computador na linguística de *corpus* revolucionaram a forma como se perquire a linguagem, tornando possível aos pesquisadores aferirem dados muito mais volumosos, e com rapidez, exatidão e riqueza de detalhes que antes eram inacessíveis. Grande parte do mérito dessa revolução provém do desenvolvimento e da aplicação de *softwares* como os concordanciadores, os alinhadores e os etiquetadores, programas de computador essenciais na tarefa do linguista de *corpus* por auxiliarem na extração das evidências empíricas que permeiem a exploração da linguagem.

Sardinha (2004) cita que o primeiro *corpus* linguístico eletrônico, o *Brown University Standard Corpus of Present-day American English*, foi lançado em 1964 e continha um milhão de palavras. Atualmente, a linguística de *corpus* está cada vez mais ligada à tecnologia e, apesar de ainda ser mais desenvolvida na Europa, no Brasil é um dos recursos que ganha cada dia mais espaço não só nas pesquisas relacionadas ao processamento da linguagem natural, lexicografia e linguística computacional, mas também nas empresas (SARDINHA, 2000).

2.3.3 Ferramentas de apoio

No campo da computação, várias ferramentas eletrônicas foram desenvolvidas a fim de facilitar o processo de coleta e análise de *corpora*. Além disso, *softwares* especializados como os concordanciadores possibilitam, a cada dia, o aprimoramento das pesquisas, promovendo automatização dos processos e obtenção de resultados mais concisos.

2.3.3.1 Concordanciadores

Conforme Tagnin (2010), concordanciadores são programas que extraem todas as ocorrências de uma palavra de busca num *corpus* juntamente com seu contexto, apresentando-as na forma de uma concordância. Essas ferramentas formam listas de concordância que agrupam todas as ocorrências das palavras ou expressões anteriormente indicadas pelo usuário. Nestas listas, a palavra buscada (que também pode ser denominada nódulo) é mostrada na posição central, em uma formatação que, na maioria dos concordanciadores, é chamada de KWIC (*Key Word In Context*). Esses programas permitem ao usuário procurar por palavras-chave em um *corpus*, fornecendo exaustivas listas da palavra em um contexto (SANTOS, 2008) e, por isso, são fundamentais para uma investigação baseada em *corpus*, especialmente quando o objetivo é estudar a colocação e a padronização lexical.

2.3.3.2 Listagem de palavras individuais (wordlist)

Outra ferramenta de grande relevância para a investigação em *corpora* é a lista de palavras individuais, também chamada *wordlist*. Trata-se de um mecanismo que cria uma lista com as palavras contidas nos textos. Essa lista é apresentada juntamente com a frequência absoluta e o percentual de cada palavra, e os termos são ordenados por ordem alfabética (que pode ser reversa, para uso em estudos de sufixos, por exemplo) ou pela frequência.

2.3.3.3 Listagem de multipalavras (clusters)

Assim como a *wordlist*, a lista de multipalavras, também conhecida como *clusters*, tem a função de produzir listas contendo todas as palavras do *corpus*. A diferença é que, nesse caso, são gerados itens compostos, ou seja, conjuntos de palavras em vez de palavras individuais (SARDINHA, 2004). As listas formadas pela ferramenta *clusters* mostram a palavra de busca juntamente com as palavras que estão agrupadas a ela, facilitando a identificação de termos como os *multi-word verbs* (Ex.: *find out*, *make up* etc.).

3. METODOLOGIA

Nesta seção serão mostrados os procedimentos adotados para a construção desta pesquisa. Em um primeiro momento serão apresentados os objetivos da pesquisa e as questões que pretendemos responder a partir desta investigação. Em seguida, trataremos dos *corpora* de estudo e de referência utilizados, bem como das ferramentas usadas na pesquisa. Por fim, mostraremos a descrição dos procedimentos utilizados para a análise dos dados.

3.1 *Objetivos e questionamentos da pesquisa*

Apesar de serem mais comuns na comunicação oral informal, os verbos frasais, juntamente com os verbos preposicionados, também ocorrem em composições formais e, dada a relevância dessas construções gramaticais, as tomamos como objeto de estudo para identificar e mapear seu uso em produções acadêmicas escritas por aprendizes brasileiros do inglês. Por isso, a partir desta investigação buscamos descobrir: a) Quais são os *multi-word verbs* mais frequentes no *corpus* de estudo? b) Esses *multi-word verbs* também ocorrem no *corpus* de referência? Com que frequência? c) De que forma os *multi-word verbs* são empregados nos textos? Existem similaridades e diferenças de uso na escrita de nativos e aprendizes brasileiros?

3.2 *Softwares utilizados na pesquisa*

Durante a realização desta pesquisa, contamos com o auxílio de diversos programas de computador, dentre eles o concordanciador *Antconc*, usado para analisar os dados linguísticos dos *corpora*; o programa de planilhas eletrônicas *Microsoft Excel*, para analisar e organizar os dados estatísticos obtidos e o *Bloco de Notas*, para transformar os dados em textos e para anular formatações prévias na transferência de dados do *Antconc* para o *Excel*.

O *Antconc* é um *software* desenvolvido pelo pesquisador britânico Laurence Anthony que permite analisar *corpora* a partir de funções como concordâncias, listas de frequências, listas de palavras individuais e múltiplas, colocações e outras. A ferramenta é *freeware*, ou seja, é um *software* gratuito, que está disponível para

download no endereço <http://www.antlab.sci.waseda.ac.jp/software.html>. No *website*, o usuário tem a opção de baixar o programa em três versões: uma para a plataforma *Windows*, uma para *Linux* e outra para *Macintosh*.

3.3 O corpus de estudo e sua composição

O *corpus* escolhido para a realização desta pesquisa é um componente brasileiro do *International Corpus of Learner English* (ICLE), intitulado Br-ICLE (*Brazilian Subcorpus of the ICLE*), formado por composições completas de alunos brasileiros que estudam a língua inglesa como habilitação na universidade. O objetivo do projeto é descrever a escrita de alunos do inglês - usando as ferramentas da linguística de *corpus* - a fim de colaborar para a tomada de decisões pedagógicas sobre ensino e aprendizagem do inglês entre brasileiros. Ainda em processo de composição, sob a responsabilidade do pesquisador Antônio Berber Sardinha, o Br-ICLE será composto por 250 mil palavras.

Para este trabalho, foi usada a versão de 2009 do Br-ICLE, que contém composições que variam de 500 a 1000 palavras cada. A partir da criação da lista de palavras, procedimento realizado com a ferramenta *wordlist* do *Antconc* para conhecer a relação de vocábulos contidos no *corpus*, obtivemos para o Br-Icle os dados de 9.575 *types* (vocábulos) e 161.204 *tokens* (ocorrências).

3.4 O corpus de referência e sua composição

Conforme citado por Sardinha (2004), o *corpus* de referência é usado para fins de contraste com o *corpus* de estudo, por isso ele não deve conter o *corpus* de estudo, pelo menos não por completo. Além disso, para que os resultados da comparação sejam significativos, é ideal que o *corpus* de referência tenha entre duas e cinco vezes o tamanho do *corpus* de estudo.

O *corpus* de referência escolhido para a realização desta pesquisa foi o LOCNESS (*Louvain Corpus of Native English Essays*), que é constituído por ensaios literários e argumentativos produzidos por estudantes nativos (britânicos e americanos) do inglês. O LOCNESS é composto por três *corpora*, e muitos dos assuntos abordados nas redações que o compõe são similares ou relacionados aos temas propostos nas composições do *corpus* de estudo Br-ICLE. O *corpus* de

referência foi utilizado para validar os padrões encontrados no *corpus* de estudo. A escolha do LOCNESS foi feita levando-se em consideração sua semelhança com o *corpus* de estudo (ambos estão em linguagem escrita, foram produzidos em meio acadêmico e são textos argumentativos).

Assim como feito com o *corpus* de estudo, os dados do *corpus* de referência também foram analisados no *software Antconc*, por meio da ferramenta *wordlist*, que nos mostrou que o LOCNESS é composto por 16.186 *types* (vocábulo) e 326.093 *tokens* (ocorrências).

3.5 Procedimentos de análise dos dados

Em síntese, os procedimentos para análise e obtenção dos termos pesquisados foram: a) Observação das estatísticas gerais do *corpus* de estudo e relação da lista de palavras na ferramenta *wordlist*, conforme mostrado nas seções 2.3.1 e 2.4.1; b) Seleção dos verbos mais frequentes no *corpus* de estudo; c) Extração dos verbos lexicais mais frequentes no *corpus* de estudo a partir da seleção de verbos elencadas no passo anterior; d) Obtenção e análise da lista de *clusters* (usando como base os verbos lexicais selecionado no passo anterior) para a extração inicial dos *multi-word verbs*; e) Obtenção e exame das linhas de concordância para a extração definitiva dos *multi-word verbs*; f) Validação dos *multi-word verbs* encontrados no *corpus* de referência; g) Comparação e análise dos *multi-word verbs* nos *corpora*.

3.6 Seleção e validação dos verbos lexicais

Após a definição dos *corpora*, o passo inicial foi a familiarização com os dados. Como descrito nos tópicos anteriores, a primeira etapa consistiu em conhecer o número de *types* (formas ou vocábulo) e *tokens* (ocorrências de palavras), bem como relacionar a lista de palavras de cada um dos *corpora*. Esse procedimento nos permitiu não só conhecer melhor os *corpora*, como também identificar os verbos mais recorrentes em cada um dos arquivos.

O segundo passo foi selecionar os verbos mais frequentes no *corpus* de estudo. Como o número de verbos é muito grande, optamos por fazer uma triagem inicial, selecionando em uma lista apenas os 46 verbos mais usados no *corpus*. A partir

daí, foi realizada uma segunda seleção, que consistiu em extrair desta lista apenas os verbos lexicais, que são os que nos interessam na pesquisa, tendo em vista a estrutura de formação dos *multi-word verbs* (verbo lexical + partícula). Nesta última triagem foram selecionados 22 verbos (em todo o processo de seleção, considerou-se somente a forma infinitiva dos verbos). Como ainda assim obtivemos um volume muito grande de palavras, decidimos analisar todos esses verbos usando a ferramenta *Clusters*. O objetivo foi verificar se as palavras que estavam mais próximas das palavras de busca eram advérbios e/ou preposições que formassem *multi-word verbs*. Assim, após a análise seria possível eliminar as palavras que não se encaixavam em nossa definição. Para realizar o procedimento configuramos a busca para que obtivéssemos *clusters* com o mínimo e o máximo de cinco palavras. Além disso, estabelecemos que o termo buscado (o verbo lexical, no caso) ficaria posicionado à esquerda, respeitando a estrutura dos *multi-word verbs* que, na maioria dos casos, é composta por verbo lexical + partículas, mas que também aceita formas com advérbios e pronomes entre o verbo lexical e as partículas (ex.: *Give it up!*).

A partir desse procedimento conseguimos eliminar verbos como *see*, *help*, *order* e *believe* que, apesar de estarem entre os mais frequentes da lista, não apresentavam formação com advérbios e preposições formando *multi-word verbs*. Para confirmar a validade desta última seleção, foi usada também a ferramenta *Concordance*. O objetivo foi observar se as palavras que eram seguidas por partículas adverbiais e/ou preposicionadas tinham realmente valor de *multi-word verbs* ou se eram apenas combinações livres⁴ entre verbos e advérbios/preposições.

Com esse procedimento, eliminamos mais alguns verbos lexicais como *like* (que fora usado, na maior parte das vezes, como preposição, e não como verbo) e *work*, que apresentou três combinações com o advérbio *out*, mas que em apenas uma delas o valor da locução correspondia a significados usuais da expressão, como “terminar bem” ou “encontrar uma solução”. Quando observadas dentro do contexto, as duas ocorrências restantes significavam “trabalhar fora de casa” e, nesse caso, uma delas constituía, possivelmente, erros da escrita acadêmica, tendo em vista a falta de um complemento ou objeto. Também foram eliminados nesta

4 Na concepção de Biber, Conrad e Leech. (2002, p. 124, tradução nossa) “Em contraste com os *multi-word verbs*, as combinações livres consistem de um verbo lexical formado por uma única palavra seguido de um advérbio. Ou preposição com um significado distinto (por exemplo, *come down*, *go back*). Na prática, é difícil fazer uma distinção absoluta entre as combinações livres e os *multi-word verbs*. É melhor pensar em um contínuo no qual algumas utilizações de verbos são relativamente idiomáticas. [...] As palavras em combinação livre retêm seus próprios significados”.

etapa os verbos que combinavam com partículas formando *multi-word verbs*, mas que ocorriam com frequência inferior a 10 (como o *make*). Assim, restaram cinco núcleos verbais: *to think, to get, to live, to take and to find*.

É importante ressaltar que todo o processo de seleção e diagnóstico dos verbos foi realizado manualmente. Cada linha de concordância foi analisada individualmente, sem o uso de ferramentas eletrônicas. A finalidade desse procedimento foi diminuir ao máximo a possibilidade de erros na classificação das palavras, já que, conforme sabemos, a formação de *multi-word verbs* pode ser realizada com advérbios e preposições e, muitas vezes, essas palavras podem desempenhar funções gramaticais diferentes em determinados contextos. Como exemplo, podemos citar o caso da palavra *up* que, de acordo com o **Concise Oxford English Dictionary** (Eleventh Edition), pode ser empregada com função de preposição, advérbio, substantivo ou mesmo verbo.

Cabe lembrar ainda que, na etapa de exame e seleção das palavras, levou-se em consideração apenas a forma infinitiva dos verbos lexicais. Durante a produção dos resultados, entretanto, as análises consideraram também a inclusão das outras formas verbais existentes para os verbos escolhidos. As formas negativas, por não serem frequentes nos *corpora*, não foram consideradas.

4. Análise e discussão dos resultados

Nesta seção, relataremos os resultados das análises realizadas com os *corpora* de estudo e de referência, bem como responderemos às questões propostas na pesquisa. Na primeira parte, serão apresentadas as estatísticas dos *corpora* utilizados. Em seguida, serão mostradas as análises dos *multi-word verbs* encontrados no *corpus* de estudo e no de referência a partir dos dados obtidos na pesquisa. Por último, trataremos da discussão dos resultados alcançados com a investigação.

4.1 Dados estatísticos dos corpora

Considerando que as grandezas numéricas dos *corpora* são diferentes, como mostrado anteriormente, fez-se necessária a normalização dos dados para definir a razão/proporção entre eles, evitando que a diferença de tamanho prejudicasse

a comparação das frequências. Para realizar o procedimento de normalização, levou-se em consideração o cálculo simples de razão que consiste na divisão de um número “a” por um número “b”. Neste caso, dividimos o valor de *tokens* do LOCNESS (*corpus* de referência) pelo do Br-ICLE (*corpus* de estudo) e obtivemos o valor de 2,02 palavras⁵, ou seja, para cada palavra do *corpus* de estudo, temos 2,02 palavras no *corpus* de referência. Além disso, a proporção dos dados foi normalizada⁶ por 100.000 palavras.

4.2 Ocorrência de multi-word verbs nos corpora

Neste tópico serão apresentadas as ocorrências de *multi-word verbs* no *corpus* de estudo com base na relação de verbos lexicais mais frequentes⁷, conforme mencionado anteriormente. Os comentários referentes à análise de cada caso encontrado seguem juntamente com os dados apresentados e são referentes às questões propostas por essa pesquisa.

4.2.1 Análise dos Multi-word verbs formados com o núcleo verbal Think

O verbo *think* é a 95ª palavra mais frequente no Br-ICLE e a 146ª no LOCNESS. Ele aparece no *corpus* de estudo 230 vezes, no entanto, se considerarmos também todas as suas formas⁸ (*thinks, thought, thinking*), esse número passa a ser de 346 ocorrências, o que mostra a relevância do verbo no universo estudado. O quadro a seguir exibe as combinações produzidas com partículas adverbiais e prepositivas a partir do verbo *to think* e suas variações:

5 Cálculo da razão dos corpora: $326.093 : 161.204 = 2.02285$.

6 A fórmula usada no cálculo para normalização dos *corpora* foi a regra de três: Frequência encontrada * 100000/Valor de *tokens* do *corpus* = x.

7 Os verbos lexicais mais frequentes foram escolhidos levando-se em consideração sua ocorrência no infinitivo e sua posição na *wordlist*. As outras formas verbais foram acrescentadas apenas na pesquisa e, mesmo que tenham contribuído para elevar a ocorrência de determinados verbos, isso não configurou critério de escolha.

8 Para nenhum dos verbos analisados foi considerada a forma negativa das palavras devido à baixa relevância no *corpus* de estudo.

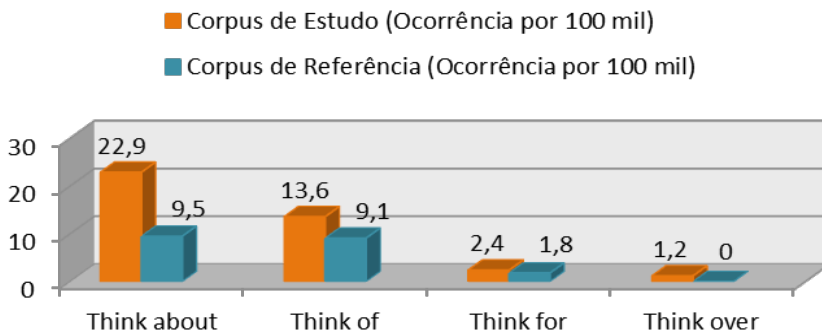
QUADRO 1- *Multi-word verbs* encontrados nos *corpora* a partir do verbo *to think*.

Núcleo verbal think nos corpora		
Multi-word verbs	Frequência (CE)	Frequência (CR)
<i>Think about</i>	37	31
<i>Think of</i>	22	30
<i>Think for</i>	4	6
<i>Think over</i>	2	0

FONTE: elaborado pelas autoras.

O verbo *to think*, no *corpus* de estudo, é encontrado formando quatro tipos de *multi-word verbs*, sendo que o mais frequente, com 37 ocorrências, é *think about*, que aparece na maioria das concordâncias analisadas com significado de “imaginar” ou “considerar” algo, como em “*In sum, it is a great utopia to **think about** [imaginar] a society in which everybody is equal [...]*” (Br-ICLE). A mesma construção lexical é também a que apresenta maior ocorrência no *corpus* de referência (31 ocorrências) e exerce a mesma função percebida em CE, como vemos em “*Then if we **think about** [considerarmos] the work day of most working people [...]*” (LOCNESS).

Com exceção do verbo frasal *think over*, todos os demais *multi-word verbs* encontrados no *corpus* de estudo ocorrem também no *corpus* de referência. No entanto, considerando-se a dimensão dos *corpora* utilizados, é possível verificar que a formação de verbos compostos por mais de uma palavra a partir de “*think*” é mais recorrente no *corpus* de estudo, conforme mostram os resultados do gráfico 1.

GRÁFICO 1: *Multi-word verbs* por ocorrência (*think*).

FONTE: elaborado pelas autoras.

4.2.2 *Multi-word verbs* formados com o núcleo verbal *Get*

O verbo *get* é a 98ª palavra mais frequente no Br-ICLE e a 135ª no LOCNESS. Ele aparece no *corpus* de estudo 216 vezes, no entanto, se considerarmos também todas as suas formas (*got*, *gotten*, *gets*, *getting*), esse número passa a ser 332, o que mostra a relevância do verbo no *corpus* de estudo. O quadro 2 exibe as combinações de *multi-word verbs* produzidas a partir do verbo *to get* e suas variações:

QUADRO 2 - *Multi-word verbs* encontrados nos *corpora* a partir do verbo *to get*.

Núcleo verbal <i>get</i> nos <i>corpora</i>		
Multi-word verbs	Frequência (CE)	Frequência (CR)
<i>Get at</i>	1	2
<i>Get in</i>	14	6
<i>Get into</i>	8	18
<i>Get from</i>	1	10
<i>Get off</i>	1	3
<i>Get on</i>	1	8
<i>Get out</i>	5	2

<i>Get out of</i>	2	10
<i>Get through</i>	4	3
<i>Get to</i>	9	30
<i>Get together</i>	9	1

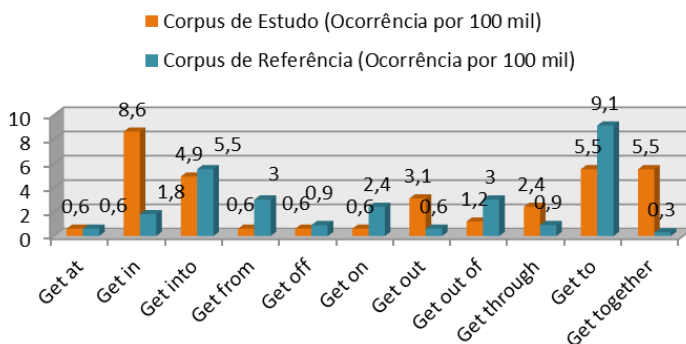
FONTE: elaborado pelas autoras.

O verbo *to get* no *corpus* de estudo é encontrado formando 11 tipos de *multi-word verbs*, sendo que, desses, a formação mais frequente, com 14 ocorrências, é *get in* (com significados de “ser admitido em algum lugar” e “entrar em algum lugar”). No *corpus* de referência, a ocorrência dessa combinação verbal é baixa, com apenas seis frequências.

Todos os *multi-word verbs* encontrados no *corpus* de estudo estão presentes também no *corpus* de referência. A comparação entre eles ressalta que, no *corpus* de nativos, a expressão *get to* é a mais frequente, com 30 ocorrências, contra apenas seis do *corpus* brasileiro. A representatividade dessa formação em CR continua maior mesmo quando observada com base nos valores normalizados. No *corpus* de referência, a locução tem função predominante de “chegar”, como em “(...) *and in an attempt to **get to** Venice, Candide loses his last sheep*” (LOCNESS). Já no *corpus* de estudo, a função varia entre “chegar” e “ter a oportunidade de fazer algo”, como em “*Unfortunately, there are people who don't **get to** learn because his/her has an external problem*” (Br-ICLE).

De todos os verbos escolhidos para análise nesta pesquisa, *to get* foi o que demonstrou o maior número de coligações lexicais formando *multi-word verbs*, no entanto, nota-se que a maior parte das expressões tem pouca frequência no universo de pesquisa. Muitos *multi-word verbs*, como *get at* e *get off*, aparecem apenas uma vez no *corpus* de estudo e com frequência igualmente baixa no *corpus* de referência.

De forma geral, e levando-se em consideração a dimensão dos *corpora* utilizados, é possível verificar que a formação de verbos compostos por mais de uma palavra a partir de “*get*” é mais usual em CR que em CE, conforme mostram os resultados do gráfico 2.

GRÁFICO 2: *Multi-word verbs* por ocorrência (*get*).

FONTE: elaborado pelas autoras.

Uma possível explicação para isso é que verbos como o *get*, que têm alta frequência, apresentam significados muito gerais quando usados sozinhos, mas bastante específicos e/ou idiomáticos quando em conjunto com outras palavras (WAIBEL, 2007), tornando-se uma armadilha para aprendizes não nativos, que em função disso recorrem ao uso de palavras equivalentes, como em “[...] *the person is never able to leave the jail [...]*” (Br-ICLE), em que *leave* exerce função similar à de *get out of* em “*They demand blacks to get out of jail [...]*” (LOCNESS).

4.2.3 *Multi-word verbs* formados com o núcleo verbal *Live*

O verbo *live* é a 100ª palavra mais frequente no Br-ICLE e a 237ª no LOCNESS. Ele aparece no *corpus* de estudo 211 vezes, no entanto, se considerarmos também todas as suas formas (*lives*, *lived* e *living*), esse número passa a ser de quase 340⁹, um número relevante no *corpus* de estudo. O quadro 3 exhibe as combinações de *multi-word verbs* produzidas a partir do verbo *to live* e suas variações:

9 Na busca por *Wordlist* no *corpus* de estudo, a palavra “*lives*” apresenta 208 ocorrências, entretanto, a maioria delas não é mostrada como verbo, mas sim como o substantivo “vidas”. Nesse caso, realizamos o cálculo das formas verbais por aproximação, inserindo na busca por *Clusters* os pronomes *he/she/it* na frente da palavra buscada.

QUADRO 3: *Multi-word verbs* encontrados nos *corpora* a partir do verbo *to live*.

Núcleo verbal <i>live</i> nos <i>corpora</i>		
Multi-word verbs	Frequência (CE)	Frequência (CR)
<i>Live by</i>	2	6
<i>Live in</i>	9	16
<i>Live for</i>	6	7
<i>Live out</i>	1	2
<i>Live together</i>	4	3
<i>Live up to</i>	1	0
<i>Live with</i>	6	15

FONTE: elaborado pelas autoras.

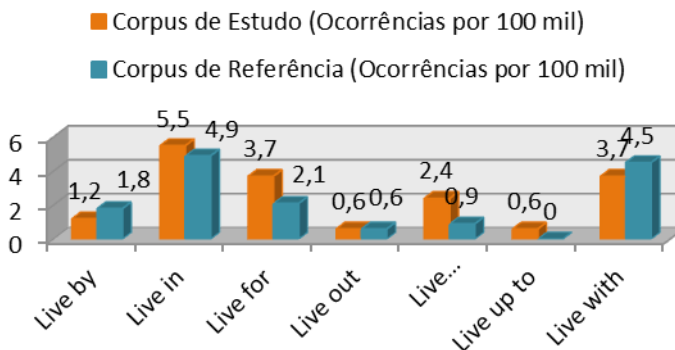
O verbo *to live* no *corpus* de estudo é encontrado formando sete tipos de *multi-word verbs*, sendo que desses, a formação mais frequente, com nove ocorrências, é *live in* (com significado de “morar no mesmo lugar em que trabalha ou estuda”). A mesma combinação também é a mais frequente no *corpus* de referência.

Ao comparar os *corpora*, podemos perceber que casos em que o verbo *to live* aparece seguido pela preposição/advérbio *in* é um dos mais recorrentes tanto em CR quanto em CE. No entanto, na maioria das concordâncias analisadas isso não configura *multi-word verbs*, já que, nestes casos, o sintagma preposicionado tem valor locativo ou de ajunto (expressando circunstância), como em “*It’s a pity that we still live in a [vivemos em uma] society that cares only about themselves*” (Br-ICLE) e “*(...) than continue to live in a [viver em um] world of contradictions, where she could not (...)*” (LOCNESS).

Outro traço que se pode perceber na comparação entre os dois *corpora* é a função que o verbo preposicionado *live with* exerce nas concordâncias analisadas. Enquanto em CR “*live with*” tem função predominante de “aceitar com desgosto” alguma situação, como em “*Therefore, the child may live with this guilt for the rest of his/her life*” (LOCNESS), em CE a função varia entre “aceitar com desgosto” e “viver com alguém sem ser casado”, como em “*There are some reasons to get married. (...) Second, it’s not so terrible to live with somebody and to accept his/her manias if you really like this person*” (Br-ICLE).

Com exceção da coligação *live up to*, todos os demais *multi-word verbs* encontrados em CE estão presentes também em CR e, de maneira geral, o uso das estruturas extraídas a partir de “*live*” é relativamente balanceado em ambos os *corpora*, conforme mostram os resultados do gráfico 3.

GRÁFICO 3: *Multi-word verbs* por ocorrência (*live*).



FONTE: elaborado pelas autoras.

4.2.4 *Multi-word verbs* formados com o núcleo verbal *Take*

O verbo *take* é a 146ª palavra mais frequente no Br-ICLE e a 127ª no LOCNESS. Ele aparece no *corpus* de estudo 153 vezes, porém, se considerarmos também todas as suas formas (*takes*, *taken*, *took* e *taking*) esse número passa a ser 282. O quadro a seguir exibe as combinações de *multi-word verbs* produzidas a partir do verbo *to take* e suas variações:

QUADRO 4 - *Multi-word verbs* encontrados nos *corpora* a partir do verbo *take*.

Núcleo verbal take nos corpora		
Multi-word verbs	Frequência (CE)	Frequência (CR)
<i>Take away</i>	10	33
<i>Take aback</i>	2	0

<i>Take on</i>	4	35
<i>Take over</i>	2	7
<i>Take up</i>	1	12

FONTE: elaborado pelas autoras.

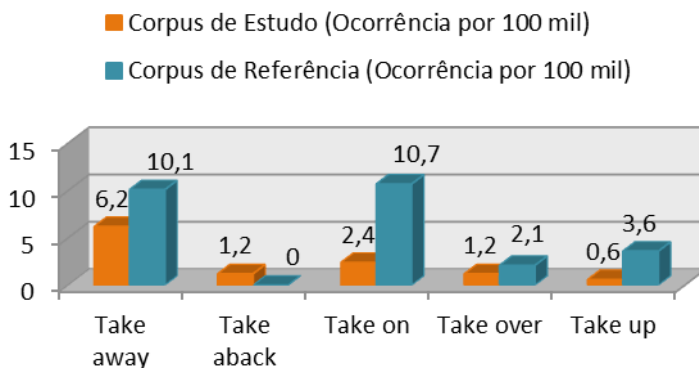
O verbo *take* no *corpus* de estudo é encontrado formando cinco tipos de *multi-word verbs*, sendo que, desses, a formação mais frequente é *take away* (na função “remover”), com 14 ocorrências. A mesma coligação lexical apresenta 33 ocorrências no *corpus* de referência, sendo a segunda mais relevante depois de *take on*.

Um dos traços que chama a atenção no estudo desse verbo é a presença da formação *take aback* (surpreender) duas vezes no *corpus* brasileiro (ex.: “*to see people simply **taken aback** by the speed and modernity*” [Br-ICLE]). Usualmente empregada como *taken aback*, a coligação não aparece nas concordâncias do *corpus* de referência. Neste último, ao contrário, surge apenas o correspondente “*be surprised*”, como em “*so that they will **not be surprised** when they get out of college straight to real*” (LOCNESS), o que pode ser um indicio da preferência dos nativos pelo uso de verbos de uma só palavra, e dos aprendizes por termos mais coloquiais (provavelmente, por não compreenderem bem os padrões do estilo argumentativo e/ou da escrita acadêmica). Ainda, apesar da pouca frequência em CE, o uso de *take aback* chama a atenção, pois não parece ser, a princípio, um *multi-word verb* muito trabalhado nas gramáticas e livros didáticos.

Outro ponto a ser destacado é a ocorrência de *take on*. A construção aparece quatro vezes no *corpus* de estudo, contra 35 do *corpus* de referência. Em sua maioria, o verbo preposicionado é usado com significado de “assumir algo”, como em “*(...) the view that Hugo joined the party to merely **take on** an identity*” (LOCNESS) e “*(...) while man made no efforts to **take on** female responsibilities*” (Br-ICLE). No entanto, percebe-se que em CE a estrutura também é usada como “levar em consideração”, como em “*Firstly if we **take on** account that lovers are artists (...)*” (Br-ICLE). Neste caso, o aprendiz utilizou um recurso pouco usual, já que os dicionários e gramáticas apresentam expressões como “*take account of*”, “*take into account*” ou “*take into consideration*”, como opções comuns entre falantes do inglês.

Com exceção da coligação *take aback*, todos os demais *multi-word verbs* encontrados em CE estão presentes também em CR e, no geral, as estruturas extraídas a partir de “*take*” são muito mais utilizadas (ou comuns) no *corpus* de referência que no de estudo, conforme mostram os resultados do gráfico 4.

GRÁFICO 4: *Multi-word verbs* por ocorrência (*take*).



FONTE: elaborado pelas autoras.

4.2.5 *Multi-word verbs* formados com o núcleo verbal *Find*

O verbo *find* é a 199ª palavra mais frequente no Br-ICLE e a 226ª no LOCNESS. Ele aparece no *corpus* de estudo 110 vezes, entretanto, se considerarmos também todas as suas formas (*found*, *finds*, *finding*), esse número passa a ser 149. O quadro a seguir exibe as combinações de *multi-word verbs* produzidas a partir do verbo *to find* e suas variações:]

QUADRO 5: *Multi-word verbs* encontrados nos *corpora* a partir do verbo *find*.

Núcleo verbal find nos corpora		
Multi-word verbs	Frequência (CE)	Frequência (CR)
<i>Find out</i>	13	25

FONTE: elaborado pelas autoras.

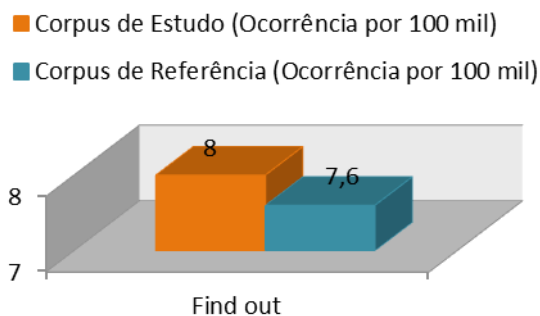
O verbo *to find* no *corpus* de estudo é encontrado formando apenas um tipo de *multi-word verbs*, o *find out*, com 13 ocorrências. A mesma estrutura também é encontrada no *corpus* de referência com 25 ocorrências.

Em ambos os *corpora* a locução é empregada com o sentido de “descobrir”, como em “*are more and more restricted to **find out** a job, even though, if this job is to guarantee (...)*” (Br-ICLE) e “*When the girls listen to their mothers they **find out** who they are.*” (LOCNESS).

A aplicação deste *multi-word verb* tanto no *corpus* de nativos, quanto no de brasileiros mostrou-se bem coerente gramaticalmente, o que sugere que esse verbo frasal é bem compreendido pelos aprendizes. Além disso, os exemplos analisados vão ao encontro da proposta de BIBER et al. (1999, p. 410), que pontua que verbos frasais transitivos, como o *find out*, são relativamente mais comuns na escrita.

De forma geral, e levando-se em consideração a dimensão dos *corpora* utilizados, é possível verificar que a formação de verbos compostos por mais de uma palavra a partir de “*find*” é equilibrada entre CR e CE, conforme mostram os resultados do gráfico 5.

GRÁFICO 5: *Multi-word verbs* por ocorrência (*find*).



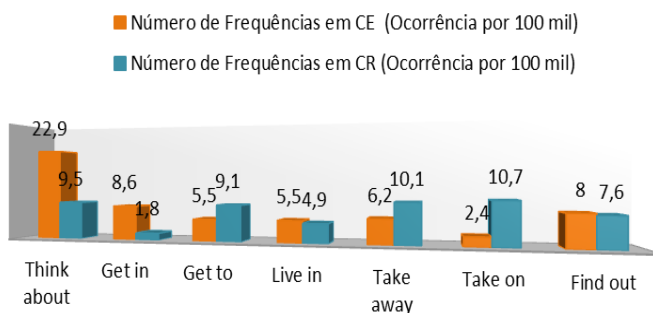
FONTE: elaborado pelas autoras.

No próximo tópico, fecharemos esta parte do trabalho com a exibição de uma síntese dos resultados apresentados e um gráfico comparativo dos verbos frasais e preposicionados mais frequentes em ambos os *corpora*. Por fim, em seguida serão apresentadas as considerações finais e as propostas para novas pesquisas.

4.3 *Multi-word verbs* mais frequentes nos *corpora*

Para finalizar esta seção, conforme análise realizada nos tópicos anteriores, o gráfico 6 fornece uma visualização decrescente dos *multi-word verbs* mais frequentes em CE e em CR para cada núcleo verbal:

GRÁFICO 6: Comparativo dos *multi-word verbs* mais frequentes em CE e em CR.



FONTE: elaborado pelas autoras.

5. Considerações finais

Este estudo buscou analisar o uso dos *multi-word verbs* nos textos acadêmicos produzidos por aprendizes brasileiros da língua inglesa, tendo como base para comparação os textos produzidos por nativos da língua. O objetivo foi identificar as unidades mais frequentes, bem como observar semelhanças e diferenças na forma como elas são empregadas em ambos os *corpora*.

Considerando os tipos de *corpora* usados (textos acadêmicos argumentativos) e também o fato de que a classe gramatical estudada (*multi-word verbs*) ocorre mais comumente em discursos orais, podemos afirmar que o número de ocorrências de verbos frasais, preposicionados e frasais preposicionados é bastante significativo em ambos os *corpora* analisados. Esse fato sugere que a hipótese de que falantes nativos utilizam mais *multi-word verbs* do que aprendizes brasileiros provavelmente é errônea, pelo menos no que diz respeito à linguagem escrita argumentativa. Para comprovar essa afirmação, entretanto, faz-se necessária uma pesquisa mais

abrangente.

É fato que, durante a análise dos dados da pesquisa nos deparamos, diversas vezes, com *multi-word verbs* iguais sendo aplicados de maneira diferente nos *corpora* de estudo e referência. Isso indica que há dúvidas em relação ao emprego dessas formas verbais em contextos mais formais, como o texto acadêmico, o que pode estar ligado às deficiências no ensino. Além disso, fica claro na pesquisa que, na escrita, os alunos brasileiros tendem, muitas vezes, a usar os *multi-word verbs* considerando o significado individual de cada palavra, o que pode ser indício, por exemplo, da influência da língua portuguesa durante a aprendizagem do segundo idioma.

Os resultados da pesquisa indicam ainda, em conformidade com diversos outros estudos disponíveis sobre o tema, que os *multi-word verbs* já não são mais um traço especificamente da linguagem falada. Ao contrário, eles têm sido amplamente utilizados também na linguagem escrita.

Cabe ressaltar aqui que esta pesquisa não tem o objetivo de ser conclusiva, mas sim de chamar a atenção para a relevância do uso de *multi-word verbs* como forma de obter proficiência na língua inglesa e, consequentemente, da necessidade de desenvolvimento de maneiras mais eficazes de ensino para essa classe gramatical. Considerando-se as limitações desta pesquisa, é certo que uma exploração mais completa sobre o tema pode ser realizada em investigações futuras, incluindo inclusive aspectos importantes relacionados ao tema, como a determinação de padrões-lexicais, usos incorretos de *multi-word verbs*, o impacto das atividades com *multi-word verbs* no aumento da fluência oral e escrita, bem como funções semânticas dessas estruturas lexicais.

A investigação aqui descrita pretende ter contribuído não só para as pesquisas envolvendo os *multi-word verbs* e suas peculiaridades no que concerne à escrita de aprendizes, mas também para a área da Linguística de *Corpus* aplicada ao ensino, por ter demonstrado a relevância dos estudos de *corpora* para a descrição de fenômenos linguísticos, a análise comportamental da linguagem e para o ensino de idiomas.

Multi-words verbs in the academic writing of english learners: a study based on a learners' *corpus*

ABSTRACT

Motivated by the importance of multi-word verbs for communication in English and also for issues surrounding the process of the EFL teaching and learning of these lexical formations, this research aims to study the use of such verbs in academic texts written by Brazilian learners of English. The aim is to identify the most common phrasal, prepositional and phrasal-prepositional verbs in the argumentative genre textual production of Brazilian learners of English, observing differences and similarities of this use to the use made by native speakers. Therefore, this study found theoretical support in Corpus Linguistics (SARDINHA, 2004, p. 3). The investigation was based on two corpora: the study *corpus*, composed by texts written by Brazilian learners of English (ICLE-Br), and another *corpus* composed of texts written by native English speakers (LOCNESS). The results suggest that Brazilian learners employ a significant amount of multi-word verbs in their argumentative texts when compared to native speakers. However, this use of multi-word verbs by Brazilian learners not always corresponds to standard patterns of the English language.

Keywords: Corpus linguistics. Multi-word verbs. Phrasal verbs. Prepositional verbs. language learning and teaching.

Referências

- BAKER, Mona. Corpora in Translation Studies: An overview and some suggestions for Future Research. **Target: International Journal of Translations Studies**. V. 7, nº 2, p. 223-243, 1995.
- BIBER, Douglas; CONRAD, Susan; LEECH, Geoffrey. **Longman Student Grammar of Spoken and Written English**. London: Pearson Education, 2002.
- BIBER, Douglas; JOHANSSON, Stig; LEECH, Geoffrey; CONRAD, Susan; FINEGAN, Edward. **Longman Grammar of Spoken and Written English**. London: Longman, 1999.
- CARTER, Ronald; MCCARTHY, Michael. **Cambridge Grammar of English**. A Comprehensive Guide: Spoken and Written English Grammar and Usage. Cambridge: Cambridge University Press, 2006.

CELCE-MURCIA, Marianne; LARSEN-FREEMAN, Diane. **The Grammar Book: An ESL/EFL Teacher's Course**. Boston, MA: Heinle and Heinle Publishing Company, 1999.

CORNELL, Alan. Realistic goals in teaching and learning phrasal verbs. **International Review of Applied Linguistics in Language Teaching**, V. 23, nº 4, pg. 269-280, 1985.

DEMPSEY, K. B.; MCCARTHY, P. M.; MACNAMARA, D. S. Using Phrasal Verbs as an Index to Distinguish Text Genres. **The Florida AI Research Society Conference - FLAIRS**, pp. 217-222, 2007. Disponível em: <<http://goo.gl/vnlqd4>>. Acesso em: 17 jan. 2016.

DOWNING, Angela; LOCKE, Philip. **A university course in English grammar**. London; New York: Routledge, 2002.

DUCROT, Oswald; TODOROV, Tzvetan. **Dicionário enciclopédico das ciências da linguagem**. São Paulo. Perspectiva, 2001.

FARIA, S. B.; BERNARDO, D. C.; SILVA, F. X. G. *Phrasal Verbs* em inglês: aprender é o melhor remédio. **Cadernos de Letras**, n. 23, p. 81-95. UFRJ, Rio de Janeiro, jan./dez. 2007. Disponível em: <<http://goo.gl/1rCKDh>>. Acesso em: 17 jan. 2016.

GREENBAUM, Sidney; QUIRK, Randolph. **A Student's Grammar of the English Language**. London: Longman, 1990.

MCENERY, Tony; WILSON, Andrew. **Corpus linguistics: An introduction**. Edinburgh: Edinburgh University Press, 2001. Disponível em: <<http://goo.gl/RygEdA>>. Acesso em: 12 jan. 2016.

PALMER, Frank Robert. **The English Verb**. London. Editora Longman Group Limited, 1974.

QUIRK, Randolph; GREENBAUM, Sidney; LEECH, Geoffrey; SVARTVIK, Jan. **A comprehensive grammar of the English language**. London: Longman, 1985.

RAMPAZZI, Elaine. **A Linguística de Corpus e os seriados Seinfeld e Friends como base para o ensino de verbos frasais e verbos preposicionados para alunos de Língua Inglesa**. 2008. 168 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem). Pontifícia Universidade Católica de São Paulo, São Paulo, 2008. Disponível em: <<http://goo.gl/Ug4hdK>>. Acesso em: 05 dez. 2015.

SANTOS, Emilson Moreira dos. **Engenharia linguística: tecnologias para apoiar as decisões gerenciais na era da Internet**. Rio de Janeiro. E-Papers, 2008.

SARDINHA, Tony Berber. **Linguística de Corpus**. Barueri, São Paulo. Editora Manole, 2004.

SARDINHA, Tony Berber. Linguística de Corpus: Histórico e Problemática. **DELTA**, São Paulo, v. 16, n. 2, 2000. Disponível em: <<http://goo.gl/lmJAJq>>. Acesso em: 23 dez. 2015.

SINCLAIR, John. **Developing linguistic corpora: a guide to good practice – Corpus and text – basic principles**. Versão de texto eletrônico. 2004. Disponível em: <<http://goo.gl/WYIYgG>>. Acesso em: 8 dez. 2015.

TAGNIN, Stella E. O. **Glossário de Linguística de Corpus**. São Paulo: HUB Editorial, 2010.

VICENTINI, Giseli P. M. **Linguística de Corpus e o seriado Friends como base para o ensino de chunks em sala de aula de Língua Inglesa**. 2006. 135 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem). Pontifícia Universidade Católica de São Paulo, São Paulo, 2006. Disponível em: <<http://goo.gl/5mw42G>>. Acesso em: 06 jan. 2016.

WAIBEL, Birgite. **Phrasal verbs in learner English: A corpus-based study of German and Italian students**. 2007. 212 f. Tese (Doutorado em Filologia). Fakultät der Albert-Ludwigs, Universität Freiburg, Alemanha, 2007. Disponível em: <<http://www.freidok.uni-freiburg.de/volltexte/3592/>>. Acesso em: 21 set. 2012.

Submetido: 30/03/2016

Aceite: 06/09/2016