

## MODOS DE NARRAR A CIÊNCIA NOS CONTEXTOS INTERNACIONAL E BRASILEIRO: PROFICIÊNCIA ESCRITA EM LÍNGUA INGLESA PARA A PRODUÇÃO DE *ABSTRACTS* EM CIÊNCIAS DA SAÚDE<sup>1</sup>

Ana Luiza Pires de Freitas<sup>2</sup>  
Maria José Bocorny Finatto<sup>3</sup>

*RESUMO:* Este artigo aborda o desenvolvimento da proficiência escrita brasileira em língua inglesa para produção de abstracts, em Ciências da Saúde. Relata-se um estudo que reuniu, descreveu e analisou um corpus de 180.170 palavras, em Medicina, Nutrição e Farmácia, à luz da Linguística de Corpus, Linguística das Linguagens Especializadas e de English for Academic Purposes. Os pacotes lexicais, sequências recorrentes de palavras empregadas nos textos, foram a unidade de análise. Com base em frequência e variabilidade lexical, constataram-se distinções nos modos de narrar a ciência entre as partes brasileira e internacional do corpus. Os achados realçam a importância do ensino a partir da ênfase nos diferentes contextos de produção de texto em saúde. Por fim, evidencia-se como os elementos da pesquisa podem ser aproveitados para criar um ambiente virtual de aprendizagem.

*PALAVRAS-CHAVE:* Proficiência escrita, Inglês especializado, Linguística de Corpus.

<sup>1</sup> Agradecemos à CAPES, ao CNPq e ao PPG-LETRAS-UFRGS pelo apoio para a realização desta pesquisa.

<sup>2</sup> Doutora em Estudos da Linguagem pela Universidade Federal do Rio Grande do Sul (UFRGS, 2016), com estágio de doutorado sanduíche na Lancaster University/UK (CAPES, 2015), Mestre em Linguagem no Contexto Social pela Universidade Federal do Rio Grande do Sul (UFRGS, 2006). Docente do Departamento de Educação e Humanidades da Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA). E-mail: analuizaf@ufcspa.edu.br

<sup>3</sup> Pós-Doutora em Ciência da Computação pelo Núcleo Interinstitucional de Linguística Computacional (NILC) do ICMC-USP (2011). Doutora em Letras (UFRGS, 2001). Professora efetiva do Depto. de Linguística, Filologia e Teoria Literária e do Programa de Pós-Graduação em Letras da UFRGS. E-mail: mariafinatto@gmail.com

*ABSTRACT: This article investigates the Brazilian development of written proficiency in English regarding the production of abstracts in Health Sciences. Based on Corpus Linguistics, Linguistics for Specialized Languages and English for Academic Purposes, the text reports a study which put together, described and analyzed a corpus of 180,170 words, in Medicine, Nutrition and Pharmacy. Lexical bundles, recurrent strings of words used in texts, were the analytical unit. From the metrics of lexical frequency and variability, the Brazilian and the international data segments uncovered distinctions in the ways of building a scientific narrative. The findings reveal the usefulness of an awareness of the different text production contexts, so that pedagogical practice can be attuned to learners' needs. To conclude with, the creation of a virtual learning environment based on the research elements is reported.*

*KEYWORDS: Written proficiency, English for Specific Purposes, Corpus Linguistics.*

*Recebido em 04/02/2017  
Aceito em 16/03/2017*

## INTRODUÇÃO

Relata-se aqui uma pesquisa que se insere no âmbito das Ciências do Léxico e toma como objeto a linguagem e as práticas textuais em Ciências da Saúde em língua inglesa. A prática textual em destaque é o texto denominado *abstract*<sup>4</sup>, que acompanha artigos acadêmicos, teses e dissertações, servindo como síntese e como um cartão de apresentação de um todo de texto. No cenário do *abstract*, foram examinadas construções recorrentes, não estritamente terminológicas, que auxiliam a caracterizar o perfil lexical desse gênero textual e

---

<sup>4</sup> Do mesmo modo como o fazem Cilvette e Pérez (2006), opta-se por utilizar o termo *abstract*, no lugar de *resumo*, que seria o correspondente em língua portuguesa, uma vez que o último é polissêmico, podendo referir-se ao texto que abre um trabalho acadêmico, interesse deste estudo, ou ao tipo de escrito que é adotado para submissão de uma pesquisa a evento acadêmico, ou mesmo ao ato de coletar e redigir as ideias principais de um texto.

discursivo<sup>5</sup>; tais construções são denominadas *pacotes lexicais* ou *lexical bundles*<sup>6</sup> (BIBER, 2006). O estudo, ao identificar, em termos quantitativos, essas construções empregadas em um acervo de *abstracts* criteriosamente reunido, propõe um aporte qualitativo, destacando os diferentes papéis pragmáticos das ocorrências no âmbito das Ciências da Saúde, especialmente Medicina, Nutrição e Farmácia. Como fruto da descrição empreendida, constata-se distinções finas entre os conjuntos de dados nacional e internacional, as quais apontam peculiaridades no modo de narrar a ciência de parte do acervo brasileiro. Este trabalho ressalta essas diferenças de construções textuais a partir das implicações para a divulgação da ciência nacional. Sustenta-se que a identificação desses aspectos pode ser vetor para uma melhor recepção da produção científica brasileira no mundo. Os achados, igualmente, reforçaram a importância de educadores linguísticos e desenhistas de programas de ensino e aprendizagem de língua inglesa reconhecerem as especificidades dos diferentes contextos e objetivos de produção textual acadêmica, para que a prática pedagógica seja sintonizada às necessidades dos aprendizes. Por fim, apresentam-se ao leitor os desdobramentos da investigação rumo à construção de um conjunto de recursos tecnológicos em ambiente digital de acesso público, com vistas a promover a sistematização didática das conclusões da pesquisa, bem como para contribuir com os avanços do campo de ensino e aprendizagem de *English for Academic Purposes*.

---

<sup>5</sup> A noção de gêneros textuais, de acordo com Ciaspuscio (2003) se origina da conceitualização de gênero discursivo na obra de Bakhtin (1997), que propôs a existência dos gêneros do discurso e os definiu como “tipos relativamente estáveis de enunciados”. No campo da Análise do Discurso e ESP, a noção foi apresentada por John Swales (1981) fortemente relacionada aos conceitos de propósito comunicativo e comunidade discursiva. Esse autor define gênero como “uma classe de eventos comunicativos compartilhada pelos membros de uma comunidade” (SWALES, 1993, p.58).

<sup>6</sup> “Sequências recorrentes de três ou mais palavras independentemente de sua idiomaticidade e de seu status estrutural” (BIBER et al., 1999, p. 990.), a exemplo de ‘*the aim of this*’, ‘*in the case of*’ e ‘*on the other hand*’, frequentemente encontrados em textos acadêmicos.

Para traçar um fio condutor entre os aspectos relacionados, este artigo segue o seguinte roteiro: primeiramente, apresentam-se os pontos de partida, o objeto de estudo e a unidade de análise da investigação; depois, explicitam-se os pressupostos teóricos que guiaram a pesquisa, seguidos da organização metodológica; posteriormente, descrevem-se os resultados obtidos pela análise dos dados, bem como as conclusões e suas implicações para o ensino e aprendizagem; por fim, expõem-se a síntese dos resultados e os desdobramentos em fase de implementação a partir da investigação.

## DOS PONTOS DE PARTIDA, OBJETO DE ESTUDO E UNIDADE DE ANÁLISE

Para chegar às questões que a investigação pretendeu responder, partiu-se de alguns pressupostos sobre escrita acadêmica e o funcionamento da linguagem, que se sintetizam nos seguintes princípios:

1. A linguagem especializada<sup>7</sup> funciona em blocos fixos (SINCLAIR, 1991) e se orienta por princípios maiores do que a livre escolha dos usuários<sup>8</sup>;
2. O reconhecimento dos membros especialistas da comunidade acadêmica funciona como fator determinante do texto ao conferir razão de ser a esse. “Essa razão formata a estrutura esquemática do discurso e influencia e delimita a

---

<sup>7</sup> É possível fazer uma extensa discussão sobre o tema do que é língua/linguagem especializada e também se esse funcionamento em blocos alcança o todo da língua, mas não é objetivo deste trabalho discutir o *modus operandi* da língua inglesa ou de qualquer outra língua como um todo. A investigação se restringe, portanto, a um diassistema específico (Ver Cosériu, 1979.).

<sup>8</sup> Por exemplo, para apresentar as conclusões de uma pesquisa, apesar de ser vernacular dizer-se ‘*we concluded that*’, de acordo com o corpus deste estudo, o usual é adotar-se ‘*the/these results suggest that*’, o que aponta para a questão de que as escolhas linguísticas adequadas são, de fato, limitadas.

escolha de conteúdo e estilo (SWALES, 1990, p. 58).” Desta forma, os objetivos comunicativos reconhecidos pelos especialistas influenciam o tema, o registro<sup>9</sup> e a estrutura do texto;

3. A frequência de ocorrências determina o padrão ou o que é entendido como característico de uma determinada área de conhecimento especializado (HOFFMANN, 2015, p. 40) ou domínio acadêmico (BIBER e CONRAD, 2009, p. 4). Essa asserção implica dizer que os textos acadêmicos não detêm características universais e podem variar situacionalmente (BIBER e CONRAD, 2009, p. 3), inclusive tendo em vista suas condições de publicação, os campos do saber e as línguas em que são redigidos. No entanto, são os traços que podemos reconhecer como mais constantes que nos apontam o que seja mais relevante para a comunidade de usuários e o que, portanto, deve ser ensinado prioritariamente, como é interesse demonstrar em relação aos textos do acervo reunido.

Com base nesses princípios, estabeleceu-se um recorte para a investigação, de que, para ser plenamente aceito em uma comunidade acadêmica ou reconhecido pelos usuários especialistas, é preciso usar com propriedade, não apenas a língua franca das publicações científicas contemporâneas, mas também as formas de dizer a ciência. Afinal, nenhum conhecimento existe de forma autônoma, mas sim organizado pela linguagem e suas formas de realização. Portanto, para atingir os propósitos de proficiência discursiva, o aprendiz brasileiro da língua inglesa para finalidades acadêmicas precisa ser capaz de construir *abstracts* que traduzam uma estrutura textual fluente. Para tanto, cada

---

<sup>9</sup> Noção que abarca escolhas léxico-gramaticais que tipificam as classes de textos, a exemplo dos textos acadêmicos, neste caso. (Ver Biber, 2006 e Biber et al., 1999).

elemento ou item lexical do texto necessita representar um valor para formar um todo de sentido. A fluência, nesse caso, está relacionada tanto a aspectos discursivos de caráter amplo, da conta da funcionalidade dos textos, quanto às escolhas léxico-gramaticais (BIBER et al., 1999). Para dar conta do postulado estabelecido, entende-se que é necessário que o professor e o estudante disponham de dados de reconhecimento dos diferentes modos de dizer e princípios de combinabilidade de elementos linguísticos dos *abstracts* do campo da saúde. Partindo dessas reflexões, o estudo priorizou as seguintes questões:

1. Que tipo de variabilidade lexical envolve palavras-tópico e elementos conexos através dos textos do *corpus* pesquisado? Quais são os elementos lexicais fixos e os variáveis?
2. Dentre as unidades lexicais associadas às palavras-tópico, quais são as significativamente mais frequentes no corpus e como se comportam funcional e estruturalmente?

Tais perguntas se organizaram a partir do texto especializado como objeto de estudo, mirando os modos de dizer, tanto através dos termos<sup>10</sup> típicos da área e suas conexões, quanto dos aspectos discursivos.

Para recortar o objeto, toma-se como inspiração o trabalho de Lothar Hoffmann, para quem o texto especializado representa “o instrumento ou o resultado de uma atividade comunicativa socioprodutiva especializada” (HOFFMANN, 1998, p. 77). Outrossim, entendemos que é possível reconhecer pontos de contato entre os postulados desse linguista germânico e a tradição dos

---

<sup>10</sup> Unidade de compreensão e de representação que funciona em modelos cognitivos e culturais, norteados pela experiência com o mundo e com seus modos de categorização (KRIEGER e FINATTO, 2004; TEMMERMAN, 1997).

estudos de Gêneros Acadêmicos, segundo os escritos de Bakhtin (1997)<sup>11</sup>, Biber (2006) e Swales (1990), pelos quais, igualmente, a investigação se orienta, pois esses se unem face a conceberem o texto como um fenômeno com sentido social.

Para entender a estrutura do texto especializado da forma como nomeado nas questões de pesquisa e chegar aos itens lexicais significativamente mais frequentes, elegeu-se, como unidade analítica, *lexical bundles* ou pacotes lexicais. Esses se referem às sequências de três ou mais palavras mais frequentes de um registro (BIBER et al., 1999, p. 990), a exemplo de ‘*the aim of this*’, ‘*in the case of*’ e ‘*on the basis of*’<sup>12</sup>. Tais grupos poli-lexicais são identificáveis empiricamente e são recuperados através de um trabalhado direcionado por corpus (*corpus-driven*), considerando-se um critério de constância e distribuição ao logo do acervo de dados. A escolha por pacotes lexicais se deu em função de serem desafiantes para o aprendiz, uma vez que funcionam como blocos de construção para a linguagem acadêmica. Sua extração é um modo objetivo e direto de relacionar unidades multi-palavras com certo nível de estabilidade.

Não obstante exista grande divergência de denominações e definições na literatura sobre combinações de unidades lexicais, a adoção do termo ‘pacote lexical’, a exemplo de Berber Sardinha (2000), como correspondente a *lexical bundle*, se deu por ser essa a expressão consagrada pelos autores com os quais igualmente nos identificamos (BIBER e CONRAD, 1999; BIBER et al., 1999; BIBER et al., 2004; CORTES, 2004; BIBER, 2006; HYLAND, 2008<sup>a</sup>, dentre os

---

<sup>11</sup> Inclui-se Bakhtin nesta classificação, porque ele, além de ter sido um precursor dos estudos de gêneros textuais, divide-os entre gêneros primários e secundários, e defende-se que os gêneros acadêmicos pertencem a esta última categoria. (Vide Bakhtin 1992, para conhecer melhor o tema.)

<sup>12</sup> Na maioria das vezes, são estruturalmente incompletos (‘*in which the*’, ‘*where there is*’), mas, podem também se apresentar na forma de estruturas canônicas (‘*as soon as*’, ‘*on the other hand*). Costumam também ser semanticamente transparentes em oposição a expressões de caráter idiomático, embora a compreensão dessas colocações estendidas possa exigir observação do contexto linguístico de ocorrência.

principais).

## DOS POSICIONAMENTO TEÓRICO INTERDISCIPLINAR

Conforme apontado acima, o estudo se estabeleceu a partir de um posicionamento interdisciplinar, ao buscar promover aproximações entre a Linguística de *Corpus* (LC), a Linguística das Linguagens Especializadas (LLE) e os Estudos em *English for Academic Purposes* (EAP). Abordamos, a seguir, os modos como os pressupostos teóricos se cruzam para dar forma à investigação.

A LC<sup>13</sup> é adotada, especialmente em função de se tratar de uma investigação empírica, que se ocupa da coleta e análise de um conjunto de dados linguísticos organizados a partir de *corpora*. Partiu-se, portanto, de um conjunto de dados que ocupam um lugar no mundo da produção científica, com base nos quais se organizam considerações sobre a configuração da linguagem. Do mesmo modo, decidiu-se que os dados seriam processados de forma automatizada. O objetivo foi de que as constatações a respeito do funcionamento da linguagem a partir dos resultados gerados não incorressem em inadequações, nas quais uma análise a olho nu poderia incidir, em função de uma maior possibilidade de erros. Assim, a pesquisa se vale da análise de registro computadorizada, um método que identifica a variação nos textos pela ocorrência de características linguísticas e que impulsionou muitos estudos sobre o gênero *abstract* em LC, a exemplo de Hyland (2004), Gledhill (2005), Hyland e Tsé (2005), e Dayrell (2011).

Também foram orientadores os fundamentos da LC, devido ao

---

<sup>13</sup> Embora muitos estudiosos defendam que a LC é apenas uma metodologia de pesquisa, argumenta-se aqui que ela é, na verdade, muito mais do que isso e que representa uma forma de aproximação da linguagem.

entendimento de língua como um sistema probabilístico (HALLIDAY, 1992). De acordo com essa noção, é possível apontar e quantificar padrões de regularidade, destacando uma correlação entre tais traços e os contextos situacionais de uso da linguagem. Nesse sentido, pode-se reconhecer que o ambiente linguístico atua sobre a co-seleção de itens lexicais. Por essa perspectiva, prevalece que a língua se orienta por um princípio idiomático, postulado por John Sinclair (1991), de acordo com o qual a seleção de itens seria guiada por padrões maiores do que simplesmente palavras. De acordo com esse princípio, portanto, as palavras não estariam, no caso, sujeitas à discriminação do escritor fluente no domínio de um gênero textual<sup>14</sup>. Ao residir no princípio idiomático, a proposta de Sinclair possibilita a expansão dos padrões fraseológicos de uso da linguagem. Observe-se que é em função da identificação com esse princípio que a unidade de análise do trabalho são os pacotes lexicais.

Partilha-se, igualmente, do aporte da LLE, a partir dos estudos em Terminologia<sup>15</sup> centrados nos textos que contêm termos (CIAPUSCIO, 2003; HOFFMANN, 2015), pois a pesquisa se ocupou da verificação do *modus dicendi* característico de textos de uma área de conhecimento especializado. Uma particularidade da comunicação entre especialistas em um tema vem a ser a precisão da informação compartilhada, que se materializa por uma linguagem especializada. Esse modo de nomear a ciência, por sua vez, remete à *linguagem especializada* que Hoffmann (2015) define como “o conjunto de todos os

---

<sup>14</sup> Com base em no corpus deste estudo, a saber, é usual dizer-se ‘*benefits and harms of ...*’ em vez de ‘*benefits and damages*’, assim como é corrente dizer-se ‘*a lower risk of ...*’ e não a ‘*smaller risk of ...*’, mesmo sendo as segundas expressões de cada par gramaticalmente aceitas e teoricamente possíveis.

<sup>15</sup> Grafada com inicial maiúscula, em oposição à ‘terminologia’, que alude a um conjunto de termos, a designação se refere ao campo de estudos interdisciplinar filiado à Linguística Aplicada, que se ocupa dos fenômenos da comunicação técnica e científica, a partir de uma abordagem prescritiva ou descritiva, conforme a linha adotada. Estudo do vocabulário científico, que toma como ponto de partida a unidade terminológica, e também, mais contemporaneamente, estudo das práticas textuais e suas convencionalidades, incluindo a comunicação entre leigos e especialistas (FINATTO, EVERS, OLIVEIRA e SILVA, 2010; FINATTO, 2011). Para maior aprofundamento das noções, vide Zilio (2010).

recursos linguísticos que são utilizados em um âmbito comunicativo, delimitado por uma especialidade, a fim de garantir a compreensão entre as pessoas que trabalham nesse âmbito” (p. 53). É devido ao recorte voltado para o texto que, por sua vez, extrapola os limites do termo, sem, no entanto, descartá-lo, que esta investigação se identifica com a LLE (HOFFMANN, 1988; FINATTO, 2004; KRIEGER, 2008; ZILIO, 2012) ou ainda com estudos em Terminologia de Perspectiva Textual (CIAPUSCIO, 2003). O aspecto frequência de classes de palavras que designam os objetos da atividade especializada também é de especial interesse para o estudo das linguagens especializadas, o que, por sua vez, estabelece um cruzamento com a LC.

Agrega-se, outrossim, o ponto de vista dos estudos em EAP, com base nos postulados de Biber e Conrad (2009) e Biber et al. (1999), dada uma preocupação particularmente pedagógica com a promoção de usuários fluentes da língua inglesa para finalidades acadêmicas, a partir de um ensino que promova autonomia e empoderamento. Por ser orientada pelos discursos de comunicação especializada, essa área de estudos aplicados se volta para os níveis de especialização dos textos, sua relação com os usuários, os níveis de expertise que possuem e os propósitos que perseguem para aprendizagem da língua inglesa nos contextos profissionais ou acadêmicos. A abordagem, a propósito, materializa pedagogicamente os estudos de LLE. Os estudos de EAP têm demonstrado, da mesma forma, aproximação com as investigações em LC, a começar pela disponibilidade de exemplos de uso autêntico da linguagem, e não de sentenças criadas para as pesquisas e, desse modo, pouco frequentes (BIBER E CONRAD, 2009). Além disso, a aplicação de princípios empíricos na área de línguas adicionais altera o foco de atenção tradicional do falante nativo para um conjunto de dados que representa o uso efetivo da língua que se deseja

descrever. Dessa maneira, transpõe-se a barreira de que o falante não nativo de uma língua estaria em desvantagem ao realizar um estudo dessa língua. Estando a investigação da linguagem calcada em dados, a necessidade de recorrer apenas a falantes que tenham o inglês como língua materna perde o sentido. De mais a mais, no caso do Brasil, muitos educadores linguísticos se sentem desvalorizados em seus conhecimentos a partir dessa dicotomia clássica entre falantes nativos e não nativos<sup>16</sup>. Portanto, a LC também é saudada pelos que se interessam por EAP por estar em condições de romper com o paradigma tradicional e fornecer aporte ao trabalho de formação de professores-pesquisadores.

## DA ORGANIZAÇÃO DOS *CORPUS* E CRITÉRIOS PARA TRATAMENTO DOS DADOS

O Corpus CISA (nomeado pela aglutinação das palavras Ciência e Saúde) é um acervo monolíngue especializado, compreendendo *abstracts* em língua inglesa das áreas de Medicina, Nutrição e Farmácia. Esse foi o nosso *corpus* de estudo, tendo sido composto por:

- a) Artigos acadêmicos publicados nos periódicos *Annals of Internal Medicine*<sup>17</sup>, *American Journal of Clinical Nutrition*<sup>18</sup> e *British Journal of Pharmacology*<sup>19</sup>;

---

<sup>16</sup> A trajetória clássica dos estudos aquisicionistas em aprendizagem de línguas estrangeiras postula o falante nativo como parâmetro para o aprendiz, sem considerar aspectos tais como quem é esse falante em termos de formação, que campos de conhecimento domina e de que registro ou dialeto é falante, dentre outros aspectos (WILLIAMS, 2006).

<sup>17</sup> <http://annals.org> (Último acesso em 24/10/2016.)

<sup>18</sup> <http://ajcn.nutrition.org> (Último acesso em 24/10/2016.)

<sup>19</sup> [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1476-5381](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1476-5381) (Último acesso em 24/10/2016.)

- b) Artigos acadêmicos publicados por pesquisadores docentes da Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), da Universidade Federal do Rio Grande do Sul (UFRGS) e da Universidade Federal Paulista (UNESP), com base no banco de dados dessas universidades;<sup>20</sup>
- c) Dissertações de mestrado e teses de doutorado de alunos dos Programas de Pós-Graduação da Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), da Universidade Federal do Rio Grande do Sul (UFRGS), da Universidade Federal Paulista (UNESP) e da Universidade Federal de Santa Catarina (UFSC), com base no banco de dados dessas universidades;
- d) Monografias de conclusão dos Cursos de Graduação de discentes das universidades referidas em ‘c’ acima.

A inclusão de *abstracts* originários de diferentes sub-áreas do campo biomédico, visou a conferir representatividade ao acervo, no segmento. O critério de seleção dos textos foi estarem disponíveis eletronicamente de forma gratuita, bem como, no caso da coleção internacional, o critério de avaliação Qualis A, pelos parâmetros da CAPES<sup>21</sup>. Já a inclusão de gêneros/registros textuais distintos, na coleção brasileira, buscou tipificar a produção tanto de especialistas quanto de autores novíços no contexto de produção acadêmica nacional.

Para a compilação do *corpus*, foi adotada a seguinte distribuição:

**Tabela (1): *Corpus* CISA – *SubCorpus* Internacional**

<i>Corpus</i>	<i>Abstracts</i>	Palavras	Tipos
---------------	------------------	----------	-------

<sup>20</sup> A Universidade Federal de Santa Catarina (UFSC), cujas produções integram o *corpus* em ‘c’ e ‘d’ abaixo, não indexa as publicações de seus pesquisadores no site.

<sup>21</sup> Avaliação de 2014.

<b>Internacional</b>			
Total	349	90098	7645
<i>American Journal of Clinical Nutrition</i>	125	32369	3666
<i>Annals of Internal Medicine</i>	114	30601	3595
<i>British Journal of Pharmacology</i>	110	27128	3825

**Tabela (2): Corpus CISA – SubCorpus Brasileiro**

<b>Corpus Brasileiro</b>	<b>Abstracts</b>	<b>Palavras</b>	<b>Tipos</b>
Total	358	90072	9556
Artigos Publicados	118	19170	3852
Teses/Dissertações	140	46147	6299
Monografias	100	24755	4399

Para chegar a 180.170 palavras, de acordo com Berber Sardinha (2004), um *corpus* de porte pequeno (de 20 a 200 mil palavras), adotou-se um critério pragmático da ordem de conveniência, defendido por Flowerdew (2004) e Swales (2002). Também analisaram-se trabalhos significativos na exploração de *abstracts* e/ou de pacotes lexicais, com coleções de proporções semelhantes ou menores do que a organizada aqui (a exemplo de BERBER SARDINHA, 2003; DAYRELL, 2009 e SÁNCHEZ HERNÁNDEZ, 2013).

O critério da quantificação dos acervos internacional e brasileiro visou que fosse possível lidar com um número de palavras equilibrado em ambos os sub-*corpora* de estudo. Além disso, buscou-se trabalhar com a mesma frequência bruta convertida para extração dos grupos multi-palavras em ambos os conjuntos de dados.

Para tratamento do *corpus*, foi escolhido o *software* AntConc<sup>22</sup> (ANTHONY, 2014). Esse é um conjunto de ferramentas de livre acesso, o que democratiza a pesquisa em LC. O AntConc permite análise de textos de forma automática e está habilitado para trabalhar com um número ilimitado de itens, desde que os arquivos sejam salvos em formato *plain text* (.txt). Para identificar as ocorrências, usou-se a função *Clusters/N-Grams*, que resgata combinações de palavras recorrentes em um *corpus*, especificando tamanho mínimo e máximo (número de palavras) por *n-grama*<sup>23</sup>, frequência mínima e dispersão.

Para o trabalho de extração dos pacotes lexicais, foi necessário estabelecer, primeiramente, alguns critérios. Para a extensão, determinou-se quatro palavras gráficas com base no princípio norteador de que os pacotes lexicais de três palavras são muito frequentes e geralmente se abrigam em sequências de quatro palavras<sup>24</sup>. Para determinar a frequência, estabeleceu-se o ponto de corte de 5 ocorrências para 180.170 palavras, porque, para *corpora* pequenos, a literatura descreve a adoção de uma métrica de frequência bruta, da ordem de entre 2 e 10 (ALTEMBERG, 1998, DE COCK, 1998). Para o fator de dispersão, tomou-se por base os postulados de Biber *et al.* (1999), Biber e Barbieri (2007) e Cortes (2004) de que uma sequência de palavras precisa ser usada em, pelo menos, de 3 a 5 textos de um *corpus* para contar como um agrupamento poli-lexical. Adotou-se então a métrica de 5 textos diferentes, valendo o critério para cada um dos dois *subcorpora*.

<sup>22</sup> Disponível em <http://www.laurenceanthony.net/software.html> (Último acesso em 24/10/2016.) .

<sup>23</sup> *N-Gramas* são agrupamentos de palavras que co-ocorrem com frequência ao longo de um texto/*corpus*, podendo conter duas, três ou mais unidades, a exemplo dos bigramas ‘*so that*’ e ‘*in fact*’ e dos trigramas ‘*in order to*’ e ‘*to start with*’. Seus elementos são imediatamente adjacentes, sem flexibilidade de posicionamento e sua semântica não-composicional nem sempre é relevante (GRIES, 2008; DUTRA e BERBER SARDINHA, 2015).

<sup>24</sup> A exemplo de ‘*as a consequence of*’ que abriga ‘*as a consequence*’.

Após a extração das ocorrências, o trabalho empírico foi o de organizar os agrupamentos de palavras a partir de uma inspeção de afinidades entre os mesmos. Para tanto, foram consideradas duas instâncias analíticas: a) as funções retóricas ou funcionais<sup>25</sup>, para as quais estabeleceu-se uma classificação própria, com base no que apontaram os dados e b) as peculiaridades léxico-gramaticais, para as quais partiu-se das categorias postuladas por Biber *et al.* (1999). O trabalho primou por entender as ocorrências, seus encaixes e distinções primeiramente. Só depois disso, estabeleceram-se as classificações.

Conforme já afirmado, a investigação promoveu um contraste do *subcorpus* brasileiro em relação às publicações dos periódicos internacionais, dado o parâmetro de qualidade estabelecido ao redor desses textos. Portanto, a pesquisa explorou os dados com vistas a analisar as diferenças de frequência dos padrões lexicais entre os as partes componentes do *corpus*.

## O QUE REVELARAM OS DADOS

Para situar o leitor sobre o que revelaram os dados, retomam-se as questões do estudo:

1. Que tipo de variabilidade lexical envolve palavras-tópico e elementos conexos<sup>26</sup> através dos textos do *corpus* pesquisado? Quais são os elementos lexicais fixos e os variáveis?
2. Dentre as unidades lexicais associadas às palavras-tópico, quais são as significativamente mais frequentes no *corpus* e como se comportam funcional

---

<sup>25</sup> Por funções retóricas ou funcionais, entende-se o significado e os propósitos da linguagem, tais aspectos visam a conferir textura ao discurso ou a organizar o texto de acordo com as situações e o contexto de uso.

<sup>26</sup> Vale destacar que, por palavras-tópico e elementos conexos, relacionam-se, respectivamente, os termos centrais que compõem os pacotes lexicais e as demais palavras que a eles se associam na composição de cada ocorrência, a exemplo de *'the objective was to'*, em que *'objective'* tem função de palavra-tópico e *'the... was to'*, de elementos conexos.

e estruturalmente?

Embora as porções internacional e a brasileira do *corpus* fossem de proporções quase idênticas (90.098 e 90.072 palavras respectivamente), o acervo brasileiro apresentou um número um pouco menor de pacotes lexicais (88 x 96). No entanto, o acervo nacional caracteriza-se por um maior número de realizações (764 x 672), conforme é possível observar no quadro que se segue<sup>27</sup>:

**Tabela (3): Ocorrências de Maior Frequência do Corpus CISA**

Subcorpus Internacional			Subcorpus Nacional		
Freq.	Disp	Ocorrências	Freq.	Disp.	Ocorrências
23	16	<i>with the use of</i>	37	36	<i>of this study was</i>
19	15	<i>was associated with a</i>	36	36	<i>this study was to</i>
14	10	<i>associated with a lower</i>	30	28	<i>the aim of this</i>
13	11	<i>had no affect on</i>	24	23	<i>aim of this study</i>
13	13	<i>the objective was to</i>	22	21	<i>as well as the</i>
12	12	<i>body mass index bmi</i>	19	16	<i>in relation to the</i>
12	12	<i>of base case analysis</i>	19	19	<i>was to evaluate the</i>
12	10	<i>on the basis of</i>	17	17	<i>study was to evaluate</i>
12	12	<i>results of base case</i>	16	16	<i>the objective of this</i>
12	8	<i>were associated with a</i>	13	12	<i>on the other hand</i>

A questão da maior abrangência dos dados brasileiros apresenta três desdobramentos analíticos significativos e relacionados:

<sup>27</sup> Em virtude da limitação de espaço, não será possível apresentar aqui todas as ocorrências extraídas do *corpus*. Os principais pacotes lexicais relacionados serão então apresentados ao longo das asserções analíticas desta seção.

1. A dispersão demonstra que os autores adotam a mesma construção repetidamente, o que revela uma maior convencionalidade dos textos. Isso significa dizer que, na redação do *abstracts*, obedecem-se a padrões linguísticos consagrados pelo uso;
2. Corroborar-se a constatação de Hyland (2008<sup>b</sup>) de que escritores menos experientes se apoiam mais em linguagem formulaica. No entanto, para chegar a tal asserção, esse autor se baseou no número absoluto de pacotes lexicais extraídos dos dados que explorou. Já neste caso, entretanto, a constatação se faz evidente por conta de maior recorrência relacionada a uma menor variedade lexical;
3. Confirma-se também uma proposição de Biber *et al.* (2004) de que não é apenas a frequência absoluta de ocorrências que revela questões que podem ser oportunas para o ensino, a partir da análise de um *corpus*. A noção de saliência, como é o caso aqui, pode também ser, igualmente, relevante. A partir desse critério, o índice de uso associações recorrentes de palavras nos dados é de 3.3% para o acervo brasileiro e de 2.9% para o acervo internacional.

Cabe discutir então se um maior apoio em linguagem formulaica seria um demérito dos textos brasileiros em relação ao padrão especializado. Para tanto, é oportuno lembrar que não há consenso acerca da relação entre autoria e grau de *expertise* de textos que fazem maior uso de colocações estendidas na literatura. Alguns estudiosos os relacionam a pesquisadores especialistas (CHEN e BAKER, 2010; SALAZAR, 2010), e outros, a novíços ou usuários menos fluentes (HYLAND, 2008<sup>a</sup>, 2008<sup>b</sup>; SALAZAR, 2011).

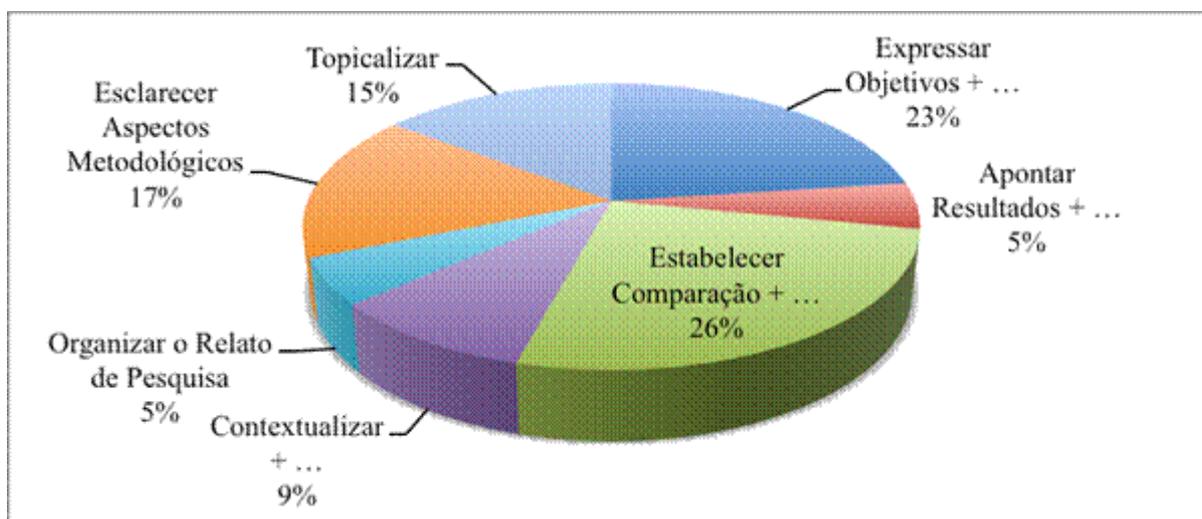
Vale, então, refletir que, uma vez que os pesquisadores brasileiros se acham fora do eixo de produção intelectual da comunidade de maior prestígio em publicações no campo da saúde, a um maior apoio em associações

recorrentes de palavras pode ser atribuído um modo de enfatizar uma relação de pertencimento ao grupo internacional de pesquisa. Além do que, a função principal dos feixes lexicais é de auxiliar os estudantes, neste caso, os usuários de inglês como língua adicional, a escreverem e falarem com maior fluência.

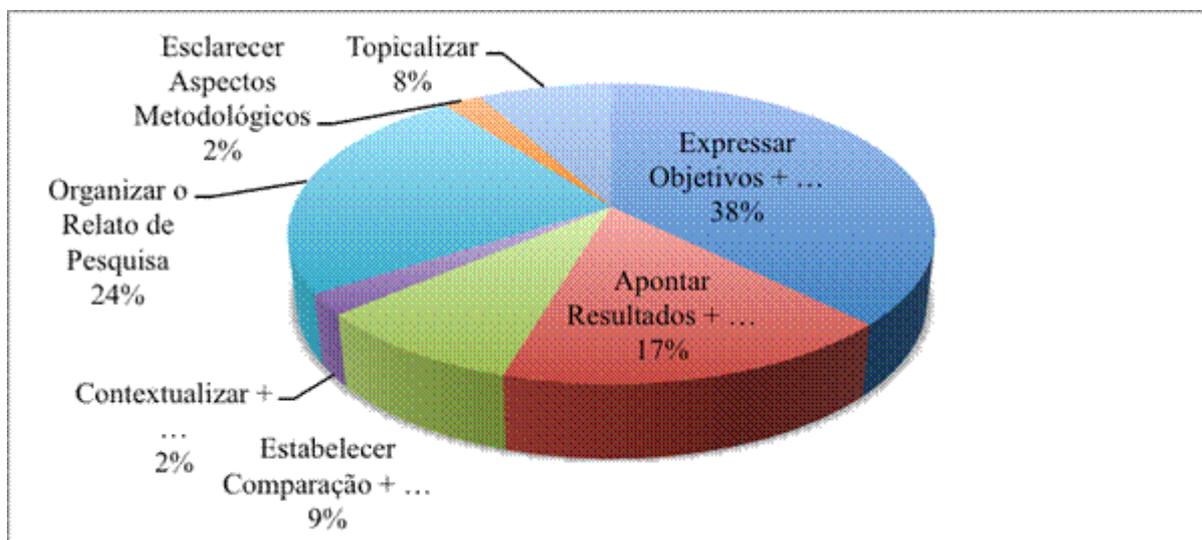
Por outro lado, é fundamental para, nos termos de Swales (1990), o pleno reconhecimento dos membros especialistas da comunidade acadêmica, que a produção textual reflita autenticidade. Por esse atributo, entende-se variedade lexical. Nesse ponto, ao ser cotejado ao acervo internacional, o *subcorpus* brasileiro se mostra menos característico do padrão de conhecimento especializado dos textos em língua inglesa, em função da baixa variedade das principais colocações estendidas implicada na repetitividade do uso de palavras tópico e elementos conexos.

Além das diferenças em relação à saliência, os acervos também se caracterizam pela presença de desigualdades funcionais e estruturais, que sugerem distinções com relação, tanto ao modo de produzir um relato científico, quanto ao uso da língua. Para ilustrar esse aspecto, apresentam-se os perfis gráficos que condensam os dados:

### **Figura (1) Distribuição Funcional do Subcorpus Internacional**



**Figura (2) Distribuição Funcional do Subcorpus Brasileiro**



A partir do estabelecimento de uma linha de corte de 10%, adotada para determinar as categorias mais representativas de cada coleção, no segmento internacional, destacam-se quatro classes retóricas. Essas contabilizam 81% das ocorrências que, para fins de síntese analítica, foram compiladas em três grupos de propósitos:

1. Fazer referência ao fazer científico propriamente: o estabelecimento de comparações +... e a indicação de aspectos metodológicos;
2. Apontar a motivação para desenvolvimento dos estudos: a expressão de objetivos +...;
3. Remeter à matéria do trabalho científico: a topicalização.

Por referência ao fazer científico, conectam-se grupos poli-lexicais tais quais *'was associated with a'*, *'associated with a lower'*, *'had no effect on'* e *'were associated with a'*<sup>28</sup>, além de *'of base case analysis'*, *'were randomly assigned to'* e *'hazards models were used'*, expressões cujas palavras-tópico e elementos conexos remontam ao inventário da atuação experimental propriamente, na proporção de 43% desta porção dos dados. Por apontamento da motivação para o desenvolvimento do trabalho acadêmico, a exemplo de *'for the treatment of'*, *'this study was used to'*, *'regression was used to'* e *'to estimate the cost'*, associam-se grupos de palavras que reiteram a motivação para a investigação científica, o que equivale a 23% das ocorrências internacionais. Em relação ao terceiro grupo, por sua vez, é possível assumir que a remissão à matéria do trabalho científico permeia as demais classificações funcionais relacionadas, especificamente, a referenciar o trabalho científico. Afinal, o segmento se categoriza pela presença de associações de palavras-tópico que especificam o conteúdo das investigações, como em *'assessment of insulin resistance'*, *'homeostasis model assessment of'*, *'mass index in kg'* e *'prospective investigation into cancer'*. Daí a importância desta última categoria funcional, especialmente, estando a mesma dentre as principais funções retóricas do acervo internacional, na proporção de 15% das ocorrências. Desse modo, nos

---

<sup>28</sup> Destaque-se a alta frequência dessas quatro expressões, que representam 15.5% das realizações do *subcorpus*.

dados internacionais, sobressai-se o fazer referência à atuação científica em si mesma.

No gráfico de dados brasileiros, por outro lado, o perfil de função retórica dos pacotes lexicais assume um contorno distinto. Destacaram-se três classes funcionais, que representam 79% das ocorrências. Nesse caso, relacionam-se as classes aferidas a, igualmente, três grupos de propósitos:

1. Apontar a motivação para desenvolvimento dos estudos: a expressão de objetivos +...;
2. Detalhar aspectos da organização das investigações: organizar o relato de pesquisa;
3. Fazer referência ao fazer científico propriamente: apontar resultados +... .

Nesse segmento, 38% dos agrupamentos recorrentes de palavras se associam para marcar as metas dos estudos (tais quais *'of this study was'*, *'this study was to'*, *'the aim of this'* e *'aim of this study'*<sup>29</sup>) e 24% desses, para organizar o relato de pesquisa (como em *'as well as the'*, *'on the other hand'*, *'in relation to the'* e *'at the end of'*). A referência à especificidade do experimento relatado propriamente dito, expressa através do apontamento dos resultados / conclusões +... (a exemplo de *'it was observed that'*, *'these results suggest that'*, *'it was concluded that'* e *'the effects of the'*), circunscreve-se ao terceiro grupo, ou seja, a 17 % dos dados, portanto.

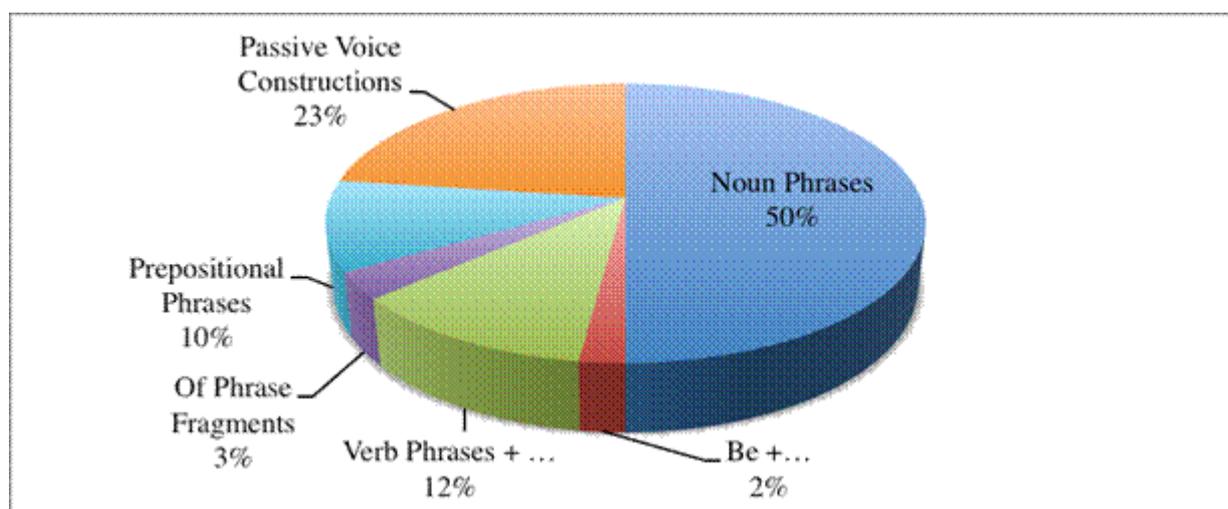
Em resumo, ao adotar os dados internacionais como padrão de referência, dado o fator de excelência que parametriza os textos, é possível perceber que os feixes de palavras desse subcorpus se voltam, centralmente, para o relato do desenvolvimento da investigação. No segundo acervo, é o apontamento da motivação para o desenvolvimento dos estudos que ocupa o percentual mais

<sup>29</sup> Esses quatro pacotes lexicais representam 16.6% das realizações dos dados brasileiros.

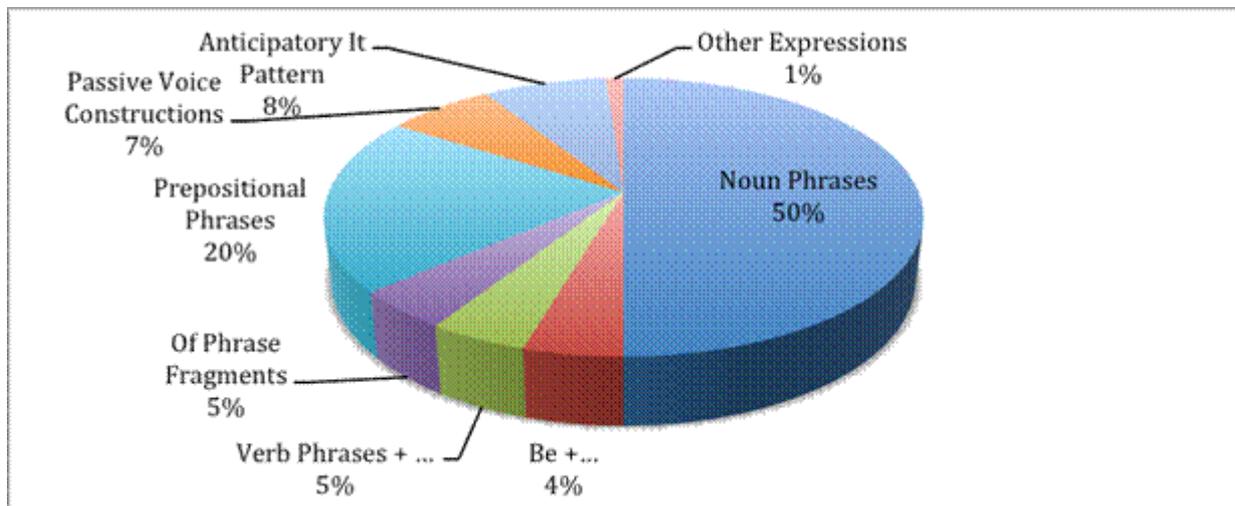
elevado. Tal noção é reforçada pelo propósito de detalhamento de aspectos da organização das investigações. A classe funcional de organização do relato de pesquisa materializa, linguisticamente, esse objetivo por meio da adoção de pacotes lexicais que trabalham para dar suporte à narrativa, especialmente, quanto à coesão textual. Dito de outro modo, no acervo internacional, os pacotes lexicais operam, principalmente, como ferramenta do fazer científico nos textos, enquanto, nos dados brasileiros, é a justificativa do trabalho acadêmico que recebe destaque.

Em sequência, apresentam-se as distribuições dos perfis léxico-gramaticais e, com isso, relacionam-se os modos como a segunda classificação dos feixes de palavras reforça a primeira:

**Figura (3): Distribuição Léxico-Gramatical do Subcorpus Internacional**



**Figura (4): Distribuição Léxico-Gramatical do Subcorpus Brasileiro**



Nos dados internacionais, salientam-se quatro categorias, que representam 95% dos feixes de palavras:

1. Noun Phrases - *'preventive services task force', 'incremental cost effectiveness ratios', 'cox proportional hazards models' e 'selection randomized controlled trials'*, dentre outros;
2. Passive Voice Constructions - tais quais *'was associated with a', 'associated with a lower', 'were associated with a' e 'were randomly assigned to'*;
3. Verb Phrases +... - a exemplo de *'had no effect on', 'to estimate the cost', 'to examine the association' e 'plays an importante role'*;
4. Prepositional Phrases – como em *'with the use of'<sup>30</sup>, 'on the basis of', 'for the treatment of' e 'in the presence of'*.

Nos dados brasileiros, são duas as categorias de destaque, as quais correspondem a 70% dos dados:

1. Noun Phrases – *'the aim of this', 'aim of this study', 'the objective of this' e 'objective of this study'* dentre outros;
2. Prepositional Phrases – *'as well as the', 'in relation to the', 'on the other hand' e 'at the end of', etc.*

<sup>30</sup> Ocorrência de maior saliência do subcorpus.

Primeiramente, ressaltam-se as categorias que se sobressaem em ambas as coleções: ‘*Noun Phrases*’ e ‘*Prepositional Phrases*’. A partir da prevalência dessas construções, estruturalmente, pode-se tomar os dois *subcorpora* como representantes típicos da prosa acadêmica, uma vez que Biber *et al.* (1999) e Hoffmann (2015) reforçam, desde longa data, que os textos especializados apresentam, em média, 60% de seu corpo composto por substantivos e adjetivos, que se sobrepõem às outras classes de palavras, com vistas a designar os objetos da atividade especializada. Entretanto, a incidência de ‘*prepositional phrases*’ no acervo brasileiro, em que a estrutura é empregada, principalmente, para organizar o relato de pesquisa, representa o dobro da proporção dessa categoria no acervo internacional. Tendo em vista que essa é a segunda categoria de maior frequência nos dados nacionais, a incidência alta poderia sugerir um estilo narrativo característico nos textos dessa coleção. Isso porque é para cumprir a função retórica de detalhar aspectos da organização das investigações que o uso de locuções preposicionadas se destaca.

Em resumo, nos dados brasileiros, as associações de palavras-tópico e elementos conexos, na forma de locuções preposicionadas, sugerem uma retórica menos objetiva do que o estilo discursivo acadêmico anglófono. Para ilustrar esse ponto, apresenta-se o quadro comparativo que correlaciona o emprego de locuções preposicionadas à função de organização do relato de pesquisa nas duas porções do acervo:

**Tabela (4): Locuções Preposicionadas na Função de Organização do Relato de Pesquisa**

Subcorpus Internacional	Frequência	Distribuição	Ocorrência
Organizar o Relato de Pesquisa	23	16	<i>with the use of</i>
	12	10	<i>on the basis of</i>
	7	5	<i>in the presence of</i>

	5	5	<i>in the absence of</i>
<b>Subcorpus Brasileiro</b>	<b>Frequência</b>	<b>Distribuição</b>	<b>Ocorrência</b>
<b>Organizar o Relato de Pesquisa</b>	22	21	<i>as well as the</i>
	19	16	<i>in relation to the</i>
	13	12	<i>on the other hand</i>
	12	7	<i>at the end of</i>
	9	8	<i>in the presence of</i>
	7	7	<i>in the field of</i>
	7	7	<i>in the state of</i>
	7	7	<i>in this study the</i>
	7	6	<i>in this study we</i>
	6	6	<i>in the city of</i>
	6	6	<i>the other hand the (frag.)</i>
	6	6	<i>in the process of</i>
	5	5	<i>in the present work* (sic)</i>
	5	5	<i>to the presence of</i>

Com base na tabela cima, tem-se 131 realizações dessa estrutura no acervo brasileiro em relação a 47 nos dados internacionais, sugerindo uma retórica mais prolixa nos textos dos autores nacionais.

Ainda acerca das categorias léxico-gramaticais em destaque, ao nos guiarmos pelo princípio de que o *corpus* é o elemento balizador, é relevante discutir as diferenças entre as demais construções mais frequentes nos dois acervos. Assim, vale notar a inclusão de ‘*passive voice constructions*’ e de ‘*verb phrases*’ como grupos principais dentre as publicações internacionais e pouco representativos no acervo brasileiro, por outro lado. Essas diferenças, novamente, sugerem um contraste no modo como a ligação entre palavras-tópico e elementos conexos, nos textos das duas coleções, reforçam a ação científica. Em relação ao uso de voz passiva, realça-se a distinção porque a estrutura é referida por pesquisadores como Biber e Conrad (2009), Swales e Feak (2009) e Hoffmann (2015) como marca do registro técnico-científico. Esse ponto parece importante porque os feixes lexicais do acervo internacional que fazem uso da construção passiva exercem, principalmente, o propósito de fazer

referência ao fazer científico. Desse modo, embora as estruturas passivas minimizem a importância do sujeito da oração e, muitas vezes, omitam o agente da sentença, o fato do foco de interesse da narrativa ser deslocado para o verbo é capaz de conferir agência ao objeto do fazer científico, como pode ser particularmente relevante para a retórica acadêmica enfatizar.

Igualmente, constata-se diferença no uso de ‘*verb phrases*’, na proporção de 12% para 5%. Com esse respeito, a menor utilização recai, novamente, sobre os dados brasileiros, em que as construções de base verbal são utilizadas somente para o propósito de expressar os objetivos +... das investigações. Nos dados internacionais, por outro lado, as combinações recorrentes de palavras iniciadas por verbos, além de expressarem a função de apontar a motivação para desenvolvimento dos estudos, também fazem referência ao fazer científico propriamente. Reiteram-se, portanto, a partir do componente léxico-gramatical, características observadas anteriormente acerca da classificação retórica das associações de palavras-tópico e seus elementos conexos.

## DA SÍNTESE DOS RESULTADOS E DOS DESDOBRAMENTOS DA PESQUISA

Em resumo, o cotejamento desenvolvido indica distinções finas entre os dois conjuntos de dados, apontando peculiaridades no modo de narrar a ciência de parte do acervo brasileiro. Essa porção do *corpus* apresenta menor variedade lexical e um maior emprego de feixes de palavras para expressar a finalidade e justificar a realização do trabalho acadêmico, empreitada que, no conjunto de dados nacionais, configura-se mais relevante do que o relato do fazer científico em si. Esse aspecto culmina por estabelecer objetivos comunicativos próprios

para os textos; tais objetivos impactam o registro acadêmico produzido, com base em uma retórica mais detalhista e menor articulação textual em promover agência ao objeto do fazer científico.

No caso dos achados descritos, produziram-se indícios significativos acerca do estilo narrativo brasileiro, no contexto pesquisado, para fomentar o ensino e, com isso, qualificar a produção acadêmica nacional. Não se trata, apenas, de homogeneizar a redação de *abstracts* brasileira no campo da saúde. Evidentemente, o texto brasileiro precisa ser léxico-gramaticalmente fluente. Mas, em termos de estrutura retórica, por outro lado, é fundamental que o ensino seja capaz de oportunizar a identificação das implicações das escolhas feitas, com relação às associações de palavras empregadas.

Dessa maneira, a partir das constatações com base no *Corpus CISA*, poder-se-ia orientar a prática linguística para incluir reflexão acerca do que é mais importante priorizar no texto que apresenta a pesquisa científica. Serão as prioridades idênticas para artigos, teses, dissertações e monografias de conclusão de curso de graduação? É possível que não, uma vez que tais escolhas se relacionam diretamente aos objetivos que cada investigação tem por meta enfatizar. De todo modo, o princípio que norteia este estudo é de que aprender a associar palavras com base no que é mais usual para produzir textos em um dado gênero escrito fornece suporte ao acadêmico do campo da saúde para a tarefa de fazer escolhas linguísticas informadas. Com isso, é possível desenvolver referências acerca do tipo de linguagem que se espera que um pesquisador produza. Contudo, é fundamental que tais evidências sejam organizadas didaticamente de modo a fazer sentido para o acadêmico, com vistas ao desenvolvimento de uma redação científica que espelhe fluência.

É, igualmente, oportuno considerar ainda até que ponto os *abstracts*

brasileiros precisam ser balizados pelos internacionais. Argumenta-se aqui que o verdadeiro exercício analítico consiste em, diante de um *corpus*, notar as diferenças e ponderar acerca dos efeitos produzidos nos textos. No que diz respeito a exemplificar a variedade de expressão e o uso de associações léxico-gramaticais, os textos internacionais do *Corpus* CISA representam um exemplo de excelência. Mas é preciso ter cautela para não incorrer em generalizações da conta de um modelo monolítico, já que os pacotes lexicais ocorrem e se comportam diferentemente em relação a ambientes disciplinares distintos.

Cumpram que educadores linguísticos e desenhistas de programas de ensino em EAP reconheçam tais aspectos, para que o ensino seja pautado pelo contexto específico de atuação do aprendiz, em oposição ao ensino de itens universalmente apropriados. Além do que, em termos dos efeitos pragmáticos que se deseja atingir em um relato científico, o que representa valor é ser conhecedor dos possíveis sentidos que o emprego das diferentes formas de associação linguística podem representar. Portanto, compete ao professor e/ou aos recursos didáticos adotado ensinar pelo exercício prático, indutivo e analítico sobre essas distinções, para que o acadêmico possa fazer escolhas informadas, a respeito das formas de expressão que dão forma ao seu fazer científico.

Por fim, cumpre destacar que, embora os pacotes lexicais mais frequentemente encontrados no *corpus* revelem aspectos gramaticais do gênero textual em foco (como em ‘*with the use of*’, ‘*of this study was*’, ‘*was associated with*’ e ‘*this study was to*’), registram-se também ocorrências com terminologias embutidas (tais quais ‘*body mass index bmi*’, ‘*preventive services task force*’, ‘*the nutritional status of*’, ‘*the family health strategy*’). Desse modo, ao deslocar-se o foco exclusivo de atenção para as terminologias empregadas nos

*abstracts*, o que, em geral, é feito em trabalhos de Terminologia sobre fraseologias especializadas, nesta pesquisa, mostra-se uma nova dimensão dos *lexical bundles*, os quais não se resumem a *um bando aleatório de unidades justapostas*. Ilumina-se assim um aspecto pouco explorado do léxico grammatical, mas que carrega também a convencionalidade do texto especializado aqui representado pelo *abstract*.

Para firmar a ponte com o objetivo geral de contribuir para o aprendizado da produção de *abstracts* em língua inglesa no campo das Ciências da Saúde, um projeto está, atualmente, em fase de desenvolvimento a partir do estudo relatado: a organização de um ambiente virtual de aprendizagem (AVA), denominado de ‘AVA Saúde’.

O propósito é dar suporte à redação de *abstracts* no segmento da saúde e apoiar, igualmente, o trabalho de sala de aula. O projeto consiste na produção de um espaço interativo para construção de conhecimento. A ideia é que o usuário tenha a oportunidade de ser co-autor, ao alimentar o conjunto de ferramentas tecnológicas com seus próprios textos, e aprendiz, ao ser levado a qualificar a escrita de modo significativo para suas necessidades acadêmicas. A adoção de um conjunto de recursos tecnológicos em ambiente digital de acesso público visa a promover a sistematização didática das conclusões da pesquisa e a contribuir com os avanços do campo de ensino e aprendizagem de EAP.

O AVA Saúde está sendo desenvolvido para comportar uma ferramenta de suporte à escrita, para apoiar a construção de *abstracts*. Através desse recurso, o autor acessará um *template* e, ao concluir seu texto, receberá um *feedback* da ferramenta, sem, no entanto, alterações automáticas no texto original, para possibilitar a verificação e comparação das sugestões feitas.

O texto ficará armazenado no sistema e alimentará a base de *feedbacks*,

formando um *corpus*. O SCIPO<sup>31</sup> (FELTRIM *et al.*, 2003), por exemplo, também é um portal de escrita acadêmica para língua inglesa disponível aos usuários brasileiros. O AVA Saúde, por sua vez, ocupará um campo de especialização em Ciências da Saúde, além de organizar os recursos pedagógicos em forma de gêneros textuais, destacando a função social dos textos no mundo dos usuários. Ademais, também se projeta abrir espaço interativo para participação dos consulentes. Para potencializar o ensino da escrita, concebemos, igualmente, além de fazer uso dos pacotes lexicais extraídos do *Corpus* CISA, compilar outros recursos, tais quais expressões formulaicas típicas dos textos acadêmicos em geral, com base em descrições a partir de outros estudos de *corpora*.

## REFERÊNCIAS

ALTENBERG, Bengt. On the Phraseology of Spoken English: The Evidence of Recurrent Word combinations. **Phraseology: Theory, Analysis and Applications**. (Ed.) COWIE, A. P. Oxford: Oxford University Press, 1998. p.101-122.

ANTHONY, Lawrence. AntConc (Version 3.2.2) [Computer Software]. Tokyo, Japan: Waseda University, 2014. Available from <http://www.antlab.sci.waseda.ac.jp/> Acesso em: out. 2016.

BAKHTIN, Mikhail. (Em Russo, 1979). **Estética da Criação Verbal**. 2a Ed. São Paulo: Martins Fontes, 1997.

BAKHTIN, Mikhail. Os gêneros do Discurso. In: **Estética da Criação Verbal**. SP: Martins Fontes, 1992. p. 277-326.

BERBER SARDINHA, Antônio Paulo. Análise de Gêneros e Linguística de **Corpus**: Identificação das unidades internas do gênero por meio de

<sup>31</sup> <http://www.nilc.icmc.usp.br/~scipo/> (Último acesso em 20/03/16.)

padronização lexical. **DIRECT Papers**, v. 51, 2003. p. 1-30.

BERBER SARDINHA, Antônio Paulo. Linguística de Corpus: Histórico e Problemática. **D.E.L.T.A.**, v.16 (2 ), 2000. p. 323-367.

BIBER, Douglas. **University language: A corpus-based study of spoken and written registers**. Amsterdam: John Benjamins, 2006.

BIBER, Douglas.; CONRAD, Susan.; CORTES, Viviana. If you look at ...: Lexical bundles in university teaching and textbooks. **Applied Linguistics**, v. 25 (3), 2004. p. 371-405.

BIBER, Douglas. JOHANSSON, Stig.; LEECH, Geoffrey.; CONRAD, Susan.; FINEGAN, Edward. **Longman grammar of spoken and written English**. Harlow: Pearson, 1999.

BIBER, D.; BARBIERE, Federica. Lexical bundles in university spoken and written registers. **English for Specific Purposes**, v. 26 (3), 2007. p. 263-286.

BIBER, Douglas; CONRAD, Susan. **Register, Genre, and Style**. Cambridge/UK: Cambridge University Press, 2009.

BIBER, Douglas; CONRAD, Susan. Lexical bundles in conversation and academic prose. In: HASSELGARD, Hilde ; OKSEFJELL, Signe (Orgs.). **Out of Corpora: Studies in Honour of Stig Johansson**. Amsterdam/Atlanta, GA: Rodopi, 1999. p.181-190.

CHEN, Yu-Hua; BAKER, Paul. Lexical Bundles in L1 and L2 Academic Writing. **Language Learning and Technology**. v.14, n.2, 2010, p 30-49.

CIAPUSCIO, Guiomar. **Textos especializados y terminologia**. Barcelona: IULA, 2003.

CILVETTI, Lurdes Divasson.; PÉREZ, Isabel Karely Leon. Textual and language flaws: Problems for Spanish doctors in producing abstracts in English. **Ibérica**, v.11, 2006. p. 61-79.

CORTES, Viviana. Lexical bundles in published and student disciplinary writing: Examples from history and biology. **English for Specific Purposes**, v. 23, 2004. p. 397–423.

COSÉRIU, Eugenio. **Teoria da Linguagem e Linguística Geral**. Rio de Janeiro/São Paulo: Presença/ EDUSP, 1979.

DAYRELL, Carmen. Anticipatory it in English abstracts: a corpus-based study of non-native student and published writing. In: GOZDZ-ROSZKOWSKI Stanislaw. (Org.) **Studies in Languages: Explorations across Languages and Corpora**. Lódz Studies in Language. 1ed. Frankfurt am Main: Peter Lang, v. 24, p. 581-598, 2011.

DAYRELL, Carmen. Lexical bundles in English abstracts: A corpus-based study of published 50 and non-native graduate writing. 2009, Liverpool. **Proceedings of the Corpus Linguistics Conference, CL2009**, 2009.

DUTRA, Deise. Prina. ; BERBER SARDINHA, Antônio Paulo. Pacotes Lexicais em *Corpora* de Aprendizes. In: IBAÑOS, Ana Maria Tramunt et al. (Orgs.) **Pesquisas e Perspectivas em Linguística de Corpus**. Porto Alegre: Mercado de letras, 2015. p. 57-80.

DE COCK, Sylvie. Repetitive phrasal chunkiness and advanced EFL speech and writing. In: MAIR, Christian; HUNDT, Marianne (Eds.). **Corpus Linguistics and Linguistic Theory**. Amsterdam: Rodopi, 2000. p. 51–68.

FELTRIM, Valéria Delisandra; ANTIQUEIRA, L. ; NUNES, Maria das Graças Volpe . A Construção de uma Ferramenta de Auxílio à Escrita de Resumos Acadêmicos em Português. 2003. In: IV Encontro Nacional de Inteligência Artificial (ENIA). Campinas. **Anais do IV ENIA**.

FINATTO, Maria José Bocorny. Estudos sobre linguagens e textos científicos e técnicos: o que é uma Terminologia Textual?. In: Elisa Battisti; Gisela Collischonn (Org.). **Língua e Linguagem: perspectivas de investigação**. 1ed. Pelotas/RS: EDUCAT, 2011. p. 153-172.

FINATTO, Maria José Bocorny. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, Aparecida Negri ; KRIEGER, Maria da Graça (Orgs.) **As ciências do léxico**:

**lexicologia, lexicografia, terminologia, volume II.** Campo Grande: Ed. UFMS/PPG-Letras UFRGS, v. 01, 2004. p. 353-363.

FINATTO, Maria José Bocorny; EVERS, Aline; ALLE, Cybele Margareth Oliveira; ALENCAR, Maria Cristina. Das terminologias às construções recorrentes: um percurso de estudos sobre linguagens especializadas. **Revista Íkala Revista de Lenguaje y Cultura**, v. 15, 2010. p. 223-258.

FLOWERDEW, Lynne. The argument for using English specialized corpora to understand academic and professional language. In: CONNOR, U.; UPTON, T. (Eds.) **Discourse in the professions: Perspectives from corpus linguistics.** Amsterdam/Philadelphia: John Benjamins, 2004. p. 11-36.

FREITAS, Ana Luiza Pires. **Proficiência Escrita em Inglês Especializado: Estudo de Corpus em Abstracts de Medicina, Nutrição e Farmácia.** Tese de doutorado. UFRGS, 2016.

FREITAS, Ana Luiza Pires ; FINATO, Maria José Bocorny. Desenvolvimento da Proficiência Escrita em Inglês Especializado: Um Estudo de *Corpus* de *Abstracts* em Ciências da Saúde. **Entrelinhas**.v.9, n.2, 2015. p. 234-266.

GLEDHILL Christopher. Problems of Adverbial Placement in Learner English and the British National Corpus. In: ALLERTON, David; TSCHIRHOLD, Cornelia ; WIESER, Judith. (Eds.) **Linguistics, Language Learning and Language Teaching. (ICSELL 10.)** Basel, Schwabe, 2005. p. 85-104.

GRIES, Stefan Th. Corpus-based methods in analyses of SLA data. In: ROBINSON, Peter ; ELLIS, Nick (Eds.), **Handbook of cognitive linguistics and second language acquisition**, 406-431. New York: Routledge, Taylor & Francis Group, 2008.

HALLIDAY, Michael Alexander Kirkwood. Corpus studies and probabilistic grammar. In: AIJMER, Karin. ; HASAN, Ruqaiya. Rationality in everyday talk: From process to system. In: SVARTVIK, J. (Org.) **Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82**, Stockholm, 4-8, August 1991. Berlin/New York: De Gruyter, 1992.

HOFFMANN, L. Conceitos Básicos da Linguística de Linguagens Especializadas. Tradução de FINATTO, Maria José Bocorny; KILIAN,

Cristiane Krause. In: FINATTO, Maria José Bocorny ; ZILIO, Leonardo (Orgs.) **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, 2015. p. 39-50.

HOFFMANN, Lothar. Grundströmungen in der Fachsprachenforschung. Traducción al catalán, Característiques dels llenguatges d'especialitat. En: HOFFMANN. Lothar. **Llenguatges d'especialitat**. Selección de textos. Barcelona, IULA, 1998.

HUCKIN, Thomas. Abstracting from abstracts. In: HEWINGS, Martin. (Ed.) **Academic writing in context: implications and applications**. Birmingham, UK: University of Birmingham Press, 2001.

HYLAND, Ken. Academic clusters: text patterning in published and postgraduate writing. **International Journal of Applied Linguistics**, v. 18, n.1, 2008<sup>a</sup>. p. 41-61.

HYLAND, Ken. Disciplinary voices. **Journal of English Text Construction** 1, v.1, Amsterdam: John Benjamins, 2008<sup>b</sup>. p. 5-22.

HYLAND, Ken. **Disciplinary Discourses: social interactions in academic writing**. Michigan: Michigan University Press, 2004.

HYLAND, Ken; TSE, Polly. Hooking the reader: a corpus study of evaluative that in abstracts. **English for Specific Purposes**, v. 24(2), 2005. p. 123-139.

ISO 214. (The International Organization for Standardization) **Documentation-Abstracts for Publications and Documentation**, 1976.

KRIEGER, Maria da Graça. Porque Lexicografia e Terminologia: relações textuais? In: FINGER, Ingrid. ; COLLISCHONN, Gisela. (Orgs.) **Anais do 8º Encontro do Círculo de Estudos Linguísticos do Sul - CELSUL**, Porto Alegre, de 29 a 31 de outubro de 2008. Pelotas: EDUCAT. 2008.

KRIEGER, Maria da Graça; FINATTO, Maria José Bocorny. **Introdução à Terminologia – Teoria e Prática**. São Paulo: Contexto, 2004.

SALAZAR, Danica. **Lexical bundles in scientific English: A corpus-based study of native and non-native writing**. Tese de doutorado. Un. de Barcelona, 2011.

SALAZAR, Danica. Lexical bundles in Philippine and British scientific English. **Philippine Journal of Linguistics**, v. 41, 2010. p. 94-104.

SÁNCHEZ HERNÁNDEZ, Purificación. Lexical bundles in three oral corpora of university students. **Nordic Journal of English Studies**, v.13, 2013. p.187-209.

SINCLAIR, John. **Corpus, Concordance, Collocation**. Oxford: OUP, 1991.

SWALES, John Malcolm. Integrated and fragmented worlds: EAP materials and corpus linguistics. In: FLOWERDEW, John Malcolm. (Ed.) **Academic discourse**. Harlow: Longman, 2002. p. 150-162.

SWALES, John Malcolm. Genre and engagement. **La Revue Belge de la Philologie et L'Histoire**, v. 71, 1993. p. 687-698.

SWALES, John Malcolm. **Genre analysis in academic and research settings**. Cambridge: Cambridge University Press, 1990.

SWALES, John Malcolm. **Aspects of article introductions**. Birmingham, UK: The University of Aston, Language Studies Unit, 1981.

SWALES, John Malcolm; FEAK, Christine. **Abstracts and the writing of abstracts**. Michigan: University of Michigan Press, 2009.

TEMMERMAN, Rita. Questioning the univocity ideal. The difference between sociocognitive Terminology and traditional Terminology. **Hermes Journal of Linguistics**, v. 18, 1997. p. 51-91.

ZILIO, Leonardo. Colocações especializadas em alemão e português na área de Cardiologia. **Tradterm**, v. 20, 2012. p. 146-177.

ZILIO, Leonardo. Terminologia Textual e Linguística de *Corpus*: estudo em parceria. In: LOPES PERNA, Cristina Becker; DELGADO KOCH, Heloísa Orsi; FINATTO, Maria José Bocorny (Orgs.). **Linguagens Especializadas em Corpora**. Porto Alegre: EDIPUCRS, 2010, p.128-151.

WILLIAMS, Geoffrey. Challenging the native-speaker norm: A corpus-driven analysis of scientific usage. In: BARNBROOK, Geoff; DANIELSSON, Pernilla; MAHLBERG, Michaela (Eds.). **Meaning Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora**. London/NY: Continuum International Publishing Group, 2006. p.115–27.