

ANA LUIZA PIRES DE FREITAS

Proficiência Escrita em Inglês Especializado:

Estudo de *Corpus de Abstracts* em Medicina, Nutrição e Farmácia

PORTO ALEGRE

2016

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ÁREA: ESTUDOS DA LINGUAGEM
ESPECIALIDADE: TEORIAS LINGUÍSTICAS DO LÉXICO

Proficiência Escrita em Inglês Especializado:

Estudo de *Corpus de Abstracts* em Medicina, Nutrição e Farmácia

ANA LUIZA PIRES DE FREITAS

ORIENTADORA: PROF^a. DR^a. MARIA JOSÉ BOCORNY FINATTO

Tese de doutorado em ESTUDOS DA LINGUAGEM/Lexicografia, Terminologia e Tradução: Relações Textuais, apresentada como requisito parcial para a obtenção do título de Doutora pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

PORTO ALEGRE

2016

CIP - Catalogação na Publicação

Freitas, Ana Luiza Pires de
Proficiência escrita em inglês especializado:
estudo de corpus de abstracts em medicina, nutrição e
farmácia / Ana Luiza Pires de Freitas. -- 2016.
254 f.

Orientadora: Maria José Bocorny Finatto.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de Pós-
Graduação em Letras, Porto Alegre, BR-RS, 2016.

1. Língua inglesa. 2. Linguística de Corpus. 3.
Proficiência escrita. 4. Ciências da Saúde. 5.
Formação de professores. I. Finatto, Maria José
Bocorny, orient. II. Título.

*Com amor: para Luiza, pela generosidade,
para Bláir, por nossa história,
para João Caetano, filho querido.*

AGRADECIMENTOS

É uma alegria concluir um trabalho da proporção de uma tese, porque representa não apenas nosso resultado intelectual, mas, igualmente, um rito de passagem acadêmico que pontua dedicação, firmeza, paciência e criatividade. Concordo com Umberto Eco, quando ele diz que “fazer uma tese significa divertir-se” (ECO, 1998, p. 169). Apesar dos momentos, não poucos, de cansaço, se não for pelo prazer, tanto do trajeto quanto da realização final, penso que não vale a pena. Por isso, destaco as pessoas que contribuíram para que esta realização tenha valido a pena.

Agradeço, especialmente e sempre, a minha orientadora querida, Maria José Bocorny Finatto, pela coragem de ter me acolhido, quando me encontrava em uma encruzilhada acadêmica. Sou grata além das palavras, pelas orientações preciosas e por suas tiradas espirituosas (‘A fé remove reservas.’ ‘Trabalhamos feitos doidas... que dupla sertaneja: Doida & Pirada.’ ‘A gente só desiste depois que morre.’ ‘Lixar é limar para ser mais objetivo.’). Que privilégio trabalhar com ela.

A minha família pelo suporte infundável e paciência com minhas ausências; coisas dos que nos amam. Ao Bláir, em especial, por ser mais rápido do que o *google* e me oferecer tantos sinônimos, ao longo da redação do texto e tantas outras coisas, que mal cabem no peito.

À Kitty, amiga maravilhosa e inspiradora, pelo carinho, erudição, papos sobre tudo e mais um pouco e por nossas adoráveis brigas. Agradeço pela leitura de parte do texto e pelo elogio de ter me considerado uma cara-de-pau na Linguística Aplicada (!). Entendi, naquele momento, que fazer tese representa assumir-se cara-de-pau.

Às amigas especiais e queridas, Simone Sarmiento e Luciene Simões, por terem me encorajado no momento certo e apontado para a Maria José, pelo apoio e o carinho mútuos que nos aproximam pela vida afora, nos momentos bons e nas horas de aperto. (*You're both something else.*)

À Marga Schlatter, amiga e incetivadora, bakhtiniana da gema, que soube me dizer palavras propulsoras, pelos momentos divertidos nas idas ao MEC e pela música deliciosa do Marmota, que compartilhamos, muitas vezes, para espantar o cansaço.

Ao meu orientador do programa de doutorado sanduíche, na Lancaster University, Tony McEnery, por sua generosidade, orientações certeiras e bem humoradas. Agradeço também à equipe do CASS, tão queridos e gentis: Kat Pardula, Carmen Dayrell, Vaclav Brezina, Dana Gablasova e Ruth Avon.

À CAPES pela bolsa concedida, que representou imensa fonte de aprendizagem. Também ao Canísio, pela eficiência na secretaria do PPG-Letras e ao pessoal da Pró-Reitoria de Pós-Graduação, pelo profissionalismo e suporte.

Ao Pedro Garcez, por ter me acolhido à entrada do doutorado. Ao Valdir Flores, coordenador do PPG à época do início do meu vínculo com a Maria José, por ter chancelado essa associação.

À Cris Killian, professora e amiga, pelos estudos e por nossas caminhadas na zona sul. À Andrea Monzón, colega inspiradora, por nossas trocas e futuras contribuições (!).

À querida Ana Cristina Ostermann, por quem tenho admiração e carinho, por nossas conversas e trocas de perto e de longe.

Ao Fabiano Piassarollo, pela paciência com sua anti-aluna, seu apoio e palavras de energia e incentivo.

Meu agradecimento também aos colegas da UFCSPA e a minha equipe do Inglês Sem Fronteiras por terem dado andamento a tudo de forma tão eficiente durante o meu período de afastamento para o sanduíche.

Por fim, aos meus alunos por me inspirarem a estudar, pesquisar, ensinar e aprender com nossas trocas.

RESUMO

Este trabalho explora o desenvolvimento da proficiência escrita em língua inglesa no âmbito da produção de *abstracts*, no campo das Ciências da Saúde. O objetivo é contribuir para a elaboração de materiais instrucionais, para a formação de educadores linguísticos e para os avanços do campo de ensino e aprendizagem de *English for Academic Purposes*. A pesquisa reuniu, descreveu e analisou um *corpus* de 180.170 palavras, com *abstracts* das áreas de Medicina, Nutrição e Farmácia, com base nos fundamentos da Linguística de *Corpus*, da Linguística das Linguagens Especializadas e dos Estudos em *English for Academic Purposes*. A unidade analítica do estudo são os pacotes lexicais (*lexical bundles*), sequências recorrentes de palavras empregadas nos textos. Para o trabalho de extração e identificação de pacotes lexicais, estabeleceu-se o critério de extensão de 4 palavras gráficas e frequência e distribuição mínimas de 5 ocorrências em, pelo menos, 5 textos diferentes, tanto para o acervo internacional, quanto para o brasileiro. Foram extraídos 96 pacotes lexicais do *subcorpus* internacional, com 90.098 palavras, e 88 sequências recorrentes do *subcorpus* brasileiro, com 90.072 palavras. Com base nas métricas de frequência e variabilidade lexical, constatam-se distinções nos modos de narrar a ciência entre as duas partes do acervo. O *subcorpus* brasileiro apresentou maior repetição de associações de palavras e um maior emprego de *lexical bundles* para expressar a finalidade e registrar a realização do trabalho acadêmico. O *subcorpus* internacional, por sua vez, caracterizou-se pela diversidade dos pacotes lexicais, pela objetividade da narrativa e pelo uso de feixes de palavras para destacar o fazer científico propriamente dito. Embora os resultados obtidos sejam específicos para o *corpus* reunido, os achados reforçam a importância de educadores linguísticos e desenhistas de programas de ensino e aprendizagem reconhecerem as peculiaridades dos contextos de produção dos *abstracts*, para que a prática pedagógica seja sintonizada às necessidades do aprendiz. Na conclusão do estudo, são apresentadas sugestões para aproveitamento dos resultados em atividades de ensino.

ABSTRACT

This research explores the development of written proficiency in English regarding the production of abstracts in the field of Health Sciences. As such, it aims at contributing to the advances in the studies of English for Academic Purposes by fostering language teachers' development, and by providing support to the creation of instructional materials. Based on Corpus Linguistics, Linguistics for Specialized Languages and English for Academic Purposes, the investigation put together, described and analyzed a corpus of 180,170 words, comprised by abstracts in Medicine, Nutrition and Pharmacy. The analytical study units are lexical bundles, recurrent strings of words used in texts. For the bundles extraction and identification, an extent criterion of 4 graphic words and a frequency and minimum distribution of 5 occurrences, in at least 5 different texts in each of the two parts of the corpus, were established. 96 lexical bundles were extracted from the international subcorpus, which adds up to 90,098 words, whilst 88 recurrent word sequences were obtained from the Brazilian subcorpus, which amounts to 90,072 words. Regarding the metrics of lexical frequency and variability, the two data segments uncovered distinctions in the ways of building up a scientific narrative. A larger repetition of word associations and a higher use of lexical bundles to express purpose and to highlight the achievement of the academic endeavor were noticed in the Brazilian subcorpus. The international subcorpus, on the other hand, features more diverse recurrent strings of words, a concise prose and the use of extended collocations to highlight the scientific enterprise in itself. Although these findings are specific to the corpus studied, they bring out the usefulness of language educators' and program designers' awareness of the peculiarities of the different abstract production contexts, so that pedagogical practice can be attuned to learners' needs. Suggestions for the application of the findings in teaching tasks are provided in the concluding part of the investigation.

ÍNDICE DE FIGURAS

FIGURA 1	Linhas de Concordância – AntiConc Versão 3.2.1	106
FIGURA 2	AntConc, Tela de Abertura	147
FIGURA 3	<i>Concordance Tool</i>	148
FIGURA 4	<i>Concordance Plot Tool</i>	148
FIGURA 5	<i>File View Tool</i>	149
FIGURA 6	<i>Clusters/N-Grams Tool</i>	149
FIGURA 7	<i>Collocates Tool</i> - Exemplo a partir do termo de busca ‘ <i>research</i> ’	150
FIGURA 8	<i>Word List Tool</i>	150
FIGURA 9	<i>Key Word List Tool</i>	151
FIGURA 10	AntConc, Caixa de Ferramentas	155
FIGURA 11	AntConc, Extensão, Frequência e Dispersão	155
FIGURA 12	AntConc, Função de Salvamento de Resultados	156
FIGURA 13	AntConc – Pacote Lexical Relacionado à Linha de Concordância	156
FIGURA 14	AntConc - Extração dos Pacotes Lexicais do SubCorpus Brasileiro de Publicações	157
FIGURA 15	AntConc - Extração dos Pacotes Lexicais do SubCorpus Brasileiro de Teses e Dissertações	158
FIGURA 16	AntConc - Extração dos Pacotes Lexicais do SubCorpus Brasileiro de Monografia	158
FIGURA 17	Gráfico da Distribuição Funcional do Subcorpus Internacional	198
FIGURA 18	Gráfico da Distribuição Funcional do Subcorpus Brasileiro	199
FIGURA 19	Gráfico da Distribuição Léxico-Gramatical do Subcorpus Internacional	202
FIGURA 20	Gráfico da Distribuição Léxico-Gramatical do Subcorpus Brasileiro	202

ÍNDICE DE TABELAS

TABELA 1	<i>Corpora</i> em Língua Inglesa Disponíveis <i>On-line</i>	101
TABELA 2	Classificação Estrutural de Pacotes Lexicais de acordo com Biber et al. (1999)	128
TABELA 3	Classificação Pragmática de Pacotes Lexicais de acordo com Biber et al. (1999)	129
TABELA 4	<i>Corpus</i> CISA – <i>SubCorpus</i> Internacional	141
TABELA 5	<i>Corpus</i> CISA – <i>SubCorpus</i> Brasileiro	141 157
TABELA 6	Padrão de Arquivamento da Identificação do <i>SubCorpus</i> Internacional	143
TABELA 7	Padrão de Arquivamento da Identificação do <i>SubCorpus</i> Brasileiro	143
TABELA 8	Exemplificando Pacotes Lexicais de 3 Palavras Abridados em Segmentos de 4 Palavras	162
TABELA 9	Ocorrências do <i>Subcorpus</i> Internacional por Classificação Funcional	163
TABELA 10	Ocorrências do <i>Subcorpus</i> Internacional por Classificação Léxico-Gramatical	172
TABELA 11	Frequência e Distribuição das Primeiras Ocorrências do <i>SubCorpus</i> Brasileiro	178
TABELA 12	Ocorrências do <i>Subcorpus</i> Brasileiro por Classificação Funcional	179
TABELA 13	Ocorrências do <i>Subcorpus</i> Brasileiro por Classificação Léxico-Gramatical	183
TABELA 14	Ocorrências do <i>Subcorpus</i> Brasileiro de Publicações	188
TABELA 15	Ocorrências do <i>Subcorpus</i> Brasileiro de Teses e Dissertações	188
TABELA 16	Ocorrências do <i>Subcorpus</i> Brasileiro de Monografias	190
TABELA 17	Palavras-Tópico e Elementos Conexos	194
TABELA 18	Ocorrências de Maior Frequência do <i>Corpus</i> CISA	195
TABELA 19	Locuções Preposicionadas na Função de Organização do Relato de Pesquisa	204
TABELA 20	<i>Corpus</i> CISA – <i>SubCorpus</i> Internacional – Pacotes Lexicais	248

TABELA 21	<i>Corpus</i> CISA – <i>SubCorpus</i> Internacional – Pacotes Lexicais	251
------------------	--	-----

ÍNDICE DE QUADROS

QUADRO 1	<i>Abstract</i> Desestruturado	75
QUADRO 2	<i>Abstract</i> Estruturado	75

ÍNDICE DE ABREVIATURAS E SIGLAS

ANSI	<i>American National Standard Institute</i>	74
AVA	Ambiente Virtual de Aprendizagem	220
CARS	<i>Create a Research Space</i>	85
CISA	Ciência e Saúde	140
EAP	<i>English for Academic Purposes</i>	20
ELT	<i>English Language Teaching</i>	20
ESP	<i>English for Specific Purposes</i>	20
IMRD	Introdução, Metodologia, Resultados e Discussão	75
IsF	Inglês sem Fronteiras	18
ISO	<i>International Standards Organization</i>	60
LA	Linguística Aplicada	21
LC	Linguística de <i>Corpus</i>	24
LLE	Linguística das Linguagens Especializadas	25
LSP	<i>Language for Specific Purposes</i>	65
MEC	Ministério da Educação e Cultura	18
PLN	Processamento de Línguas Naturais	80
TCT	Teoria Comunicativa da Terminologia	60
TGT	Teoria Geral da Terminologia	58
TC	Terminologia Cultural	61
TS	Terminologia Sociocognitivista	61

SUMÁRIO

INTRODUÇÃO: ‘THE AIM OF THIS...’	17
Objetivos	17
Motivação de Pesquisa	20
Pontos de Partida e Questões de Pesquisa	21
Objetivo de Estudo e Unidade de Análise	24
Embasamento Teórico	25
Organização do Trabalho	29
I. PRIMEIRA PARTE – A FUNDAMENTAÇÃO TEÓRICA	31
1. CAPÍTULO 1	32
TEXTO, O TECIDO DA INTERAÇÃO	33
1.1. Texto, <i>textum</i> , tecido	34
1.2. Classes de eventos relativamente estáveis e situacionais	40
1.2.1. Das contribuições de Mikhail Bakhtin	41
1.2.2. Das contribuições de John B. Swales	45
1.2.3. Das contribuições de Douglas Biber a um ponto de confluência	50
2. CAPÍTULO 2	56
O TEXTO ESPECIALIZADO COMO FORMA DE PARTICIPAÇÃO EM UM MUNDO QUE SE COMPLEXIFICA	57
2.1. Jovem ciência do léxico	57
2.2. Dando forma ao léxico da ciência	64
3. CAPÍTULO 3	72
ABSTRACTS, UMA VIRADA NA HISTÓRIA DOS TEXTOS ACADÊMICOS	73
3.1. Das origens e dos objetivos	73
3.2. Das caracterizações e do posicionamento das investigações	78
3.2.1. Do legado dos primeiros estudos	80
3.2.2. Dos estudos mais recentes que se relacionam com esta pesquisa	87
4. CAPÍTULO 4	91

A LINGUÍSTICA DE <i>CORPUS</i>, OS ESTUDOS EM <i>ENGLISH FOR ACADEMIC PURPOSES (EAP)</i> E OS PACOTES LEXICAIS (<i>LEXICAL BUNDLES</i>)	92
4.1. A Contracultura da Linguística de <i>Corpus</i>	92
4.1.1. Das concepções centrais	94
4.1.2. Do trabalho de <i>corpus</i>	98
4.1.3. Das análises computadorizadas e do principal objetivo das investigações: ‘diga-me com quem andas’	104
4.1.4. Dos tipos de estudos em LC	108
4.1.5. O estatuto do campo	111
4.2. <i>EAP</i> : o lugar das práticas comunicativas nos contextos acadêmicos	113
4.3. Desempacotando o léxico dos textos	118
4.3.1. Da escolha por pacotes lexicais (ou <i>lexical bundles</i>)	121
4.3.2. Das principais características dos pacotes lexicais	122
4.3.3. Das classificações dos pacotes lexicais	126
4.3.4. Das perspectivas das investigações	130
II. SEGUNDA PARTE – A APLICAÇÃO DO <i>CORPUS</i>	134
5. CAPÍTULO 5	135
METODOLOGIA	136
5.1. Dos das perguntas de pesquisa	136
5.2. Do <i>corpus</i> do estudo	139
5.3. Da concepção teórico-analítica	144
5.4. Das etapas da análise	151
5.4.1. Dos critérios de identificação e tratamento dos pacotes lexicais	152
5.4.2. Do cotejamento dos <i>subcorpora</i>	153
5.4.3. Do passo a passo analítico	154
6. CAPÍTULO 6	160
‘<i>THESE DATA SUGGEST THAT...</i>’: O QUE REVELAM OS DADOS...	161
6.1. Do <i>subcorpus</i> internacional	161
6.2. Do <i>subcorpus</i> brasileiro	178
7. CAPÍTULO 7	193
‘<i>THESE RESULTS SUGGEST THAT...</i>’: SÍNTESE DOS RESULTADOS	

E RETOMADA DAS QUESTÕES DE PESQUISA	194
7.1. Do tipo de variabilidade lexical e do comportamento estrutural funcional dos dados que integram o <i>corpus</i>	195
7.2. Da síntese dos resultados	206
8. CAPÍTULO 8	211
<i>‘STUDIES ARE NEEDED TO...’</i>	212
8.1. Do modo como fragmentos do léxico conferem padrões aos <i>abstracts</i>	213
8.2. Das perspectivas futuras e do alcance da pesquisa	220
9. REFERÊNCIAS BIBLIOGRÁFICAS	225
10. ANEXOS	248

“Sempre existiram drogas mais potentes, mais calmantes, mais tranquilizantes, mais alucinógenas do que todas as drogas da farmacopeia antiga e da farmacologia moderna. Essas *miracle-drugs*, essas drogas-milagre são as palavras.”

D. Segrè, 1921, p. 73

INTRODUÇÃO: ‘THE AIM OF THIS...’

Esta pesquisa, desenvolvida na linha de pesquisa *Lexicografia, Terminologia e Tradução: Relações Textuais*, do PPG-Letras da UFRGS, insere-se no âmbito das Ciências do Léxico e toma como objeto a linguagem e as práticas textuais das Ciências da Saúde em língua inglesa.

A prática textual em destaque é o texto denominado *abstract*¹, 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*, são examinadas construções recorrentes, não estritamente terminológicas, que auxiliam a caracterizar o perfil lexical desse gênero² textual e discursivo; tais construções são denominadas "pacotes lexicais". A pesquisa, 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.

O trabalho se insere no âmbito da Linguística Aplicada, buscando subsidiar o ensino de língua inglesa para propósitos específicos. Como fruto da descrição empreendida, a investigação traz dados úteis para elaboração de materiais didáticos em formato *on-line* associados ao *corpus* e às construções recorrentes identificadas.

Objetivos

O ensino superior brasileiro se encontra em época de ebulição em função dos esforços em prol da internacionalização, tais quais acordos de pesquisa e programas de mobilidade acadêmica. As universidades parceiras do exterior, por sua vez, vêm adotando,

¹ 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.

² Gêneros – tipos estáveis de textos (BAKHTIN, 1997); “padrões convencionalmente aceitos para ações linguísticas complexas” (BRINKER, 1985, p. 124). A noção de gêneros textuais, de acordo com Ciapuscio (2003), origina-se 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 *English for Specific Purposes* (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).

majoritariamente, a língua inglesa para ensino e pesquisa, mesmo em países cuja primeira língua não seja anglófona. A necessidade de garantir o domínio do idioma no Brasil se torna, portanto, cada vez mais relevante. A exemplo, o Ministério da Educação e Cultura (MEC) lançou, em julho de 2012, o Programa Inglês Sem Fronteiras (IsF), com vistas a alavancar, emergencialmente, a capacitação linguística dos alunos com o desafio de estudar em centros de ensino que adotam a língua inglesa. Além disso, promover um bom conhecimento da língua, para que os cientistas brasileiros possam publicar no idioma inglês e, com isso, fazer parte da comunidade internacional de pesquisa, se torna fator de inquestionável importância.

A divulgação da pesquisa científica, por meio de periódicos elegíveis, oportuniza tanto o reconhecimento quanto a credibilidade da ciência, duas premissas necessárias para o compartilhamento do conhecimento entre os membros da comunidade pesquisadora. É a partir do acesso à informação existente que os participantes vão ampliando sua gama de produções e, conseqüentemente, expandindo a comunicação especializada. A ciência, por seu turno, ganha forma através de estudos que se organizam e circulam através de registros em gêneros de escrita acadêmica. A comunidade científica, como grupo organizado, regula a sua atividade e o seu discurso³, cujas normas são seguidas por seus membros e reconhecidas na gama de textos utilizados para comunicar sua produção (SWALES, 1997), tais quais artigos acadêmicos, teses, dissertações e monografias, dentre outros. Pode-se dizer mesmo que as formas de dizer a ciência também são e constituem a ciência. Daí a importância de compreendê-las e adotá-las adequadamente para quem desejar ser aceito como membro da comunidade pesquisadora, especialmente, quando o acesso acontece através de uma língua adicional⁴ como a língua inglesa, *lingua franca* de muitos campos de conhecimento, a exemplo das Ciências da Saúde. Afinal, escritores competentes no contexto acadêmico, participantes de forma ativa e crítica nos eventos internacionais, terão melhores condições de tomar parte na rede de comunicação e ocupar espaço na comunidade científica internacional.

O mundo do conhecimento especializado, por outro lado, culmina por sofrer uma rígida formalização nas últimas décadas, por conta, especialmente, das exigências do ensino superior e das revistas especializadas, com base nas demandas dos gêneros textuais

³ Esta pesquisa segue a linha de Cabré (2002), Ciapuscio & Kuguel (2002) e Guimarães (2006) no que tange a não estabelecer diferenças entre texto e discurso.

⁴ O termo *língua adicional* refere-se à aprendizagem e ao uso de qualquer língua, outra que não a materna, que promove o acréscimo de práticas sociais. A adoção desse conceito reflete a necessidade de se entender a língua como um artefato cognitivo, cultural e social (HELLERMAN, 2008), em oposição à visão cognitivista tradicional que gera a dicotomia *língua estrangeira* versus *segunda língua*. Embora reconheça-se que o termo *língua estrangeira* ainda seja amplamente adotado nas pesquisas em Linguística Aplicada, ao longo desta investigação, serão respeitadas as referências às fontes citadas, mas adotada a denominação *língua adicional* para as considerações de autoria própria.

relacionados a esses contextos. Ou seja, “a produção exemplar de importantes especialistas, também conduziu à elaboração de planos de construção textual relativamente fixos” (HOFFMANN, 2015, p. 28). É nessa conjuntura que surgem os textos de interesse para este estudo, quais sejam, os *abstracts* dos textos acadêmicos.

Síntese do texto acadêmico, seja de um artigo, de uma tese ou dissertação, o *abstract* começa a ser incorporado nos anos 60 e, contemporaneamente, faz-se presente em todas as publicações de renome da área da saúde (CILVETTI e PÉREZ, 2006). Atualmente, os *abstracts* funcionam quase como um *cartão de visitas* dos textos que precedem, pois, mesmo em revistas especializadas de acesso restrito, são de domínio democratizado⁵. Sua função se destaca por permitir que o leitor acesse de modo sintético o conteúdo do trabalho, o que também vale para comitês avaliadores de eventos científicos ou para análises preliminares de publicações (SWALES e FEAK, 2009). Mesmo para pesquisas desenvolvidas em língua portuguesa, é cada vez mais usual que os *abstracts* sejam redigidos em língua inglesa. Em função do propósito de “sumarizar, indicar e predizer o conteúdo e a estrutura do texto integral que segue” (MOTTA-ROTH e HENGES, 2013, p. 152), mesmo quando o trabalho completo está disponibilizado para consulta, é comum ler-se primeiramente o *abstract*, que guia o leitor através do texto, além de influir e facilitar suas escolhas de pesquisa mais aprofundadas. Em outras palavras, a importância do *abstract* reside em sua função de orientação e de delimitação de pesquisa para uso acadêmico em geral.

A produção de *abstracts* pode parecer tarefa simples, em função de se desenvolver, na maioria das vezes, em apenas um parágrafo, mas, na verdade, é uma atividade árdua, mesmo para pesquisadores experientes, em razão dessa categoria de texto apresentar alta densidade de informação em poucos caracteres. Ainda, as formas de redigi-los variam de área para área, e o modo de produzir textos científicos em língua inglesa não é o mesmo da língua portuguesa, estando implicadas tanto questões de ordem de escolhas lexicais, quanto do âmbito retórico. Por essas razões, torna-se significativo investigá-los.

Portanto, tendo como contexto a internacionalização dos acadêmicos brasileiros e, com isso, a expansão da comunicação especializada, o tema desta investigação é o ensino da produção de *abstracts* em língua inglesa, a partir de textos acadêmicos do campo de conhecimentos da saúde. O objetivo é descrever e analisar um *corpus*⁶, compilado a partir da área especializada em que os textos se realizam (especificamente, de Medicina, Nutrição e

⁵ Grande parte dos periódicos especializados da área da saúde são acessíveis somente sob assinatura. Um exemplo de referência mundial nesse macro gênero são as publicações do *Chemical Abstracts Service*: <https://cas.silkroad.com> (Último acesso em 08/10/2015).

⁶ Conjunto de dados criteriosamente selecionados como matéria-prima para um estudo.

Farmácia) e discutir as implicações dos aspectos examinados para o ensino. Desse modo, pela caracterização de um conjunto de textos que tipifica a linguagem das Ciências da Saúde, pretende-se contribuir para produções futuras de materiais instrucionais, assim como para a formação educadores linguísticos e para os avanços do campo de ensino e aprendizagem de *English for Academic Purposes (EAP)*⁷.

Ao prosseguir com este capítulo introdutório, contextualiza-se a razão do recorte temático da pesquisa a partir da motivação vinculada a uma trajetória profissional.

Motivação de Pesquisa

No começo da década de noventa, a atuação na docência da língua inglesa em um curso-livre⁸ culminou com minha aproximação do ensino da modalidade de *English for Specific Purposes (ESP)*⁹, então na forma de inglês para negócios. Some-se a isso o fato de que o trabalho no campo de formação de educadores linguísticos traria contato com uma publicação de vanguarda para a época: *The Lexical Approach: the state of ELT and a way forward* (LEWIS, 1993). A obra chamou muito a atenção por introduzir a noção do léxico gramaticalizado e não da gramática lexicalizada. Aquele período era o auge da Abordagem Comunicativa¹⁰ para o ensino e aprendizagem de línguas adicionais, e destacar a centralidade do léxico, no sentido de palavras agrupadas para criar um texto contínuo e coerente, dava mostras de fazer todo o sentido, na medida em que a Abordagem Lexical trazia a proposta de se ensinar uma língua autêntica, já que devia ser pautada pelo uso.

Surgiu então a oportunidade de organizar materiais de ensino em *ESP* para uma

⁷ O campo de *English for Academic Purposes (EAP)*, subcampo de *English for Specific Purposes (ESP)*, nasceu nas últimas décadas do séc. XX, voltado para a descrição dos gêneros/registros textuais acadêmicos, com vistas à produção de material pedagógico, bem como para o ensino/aprendizagem de língua, a partir das necessidades de estudantes e pesquisadores que precisam dominar a língua inglesa (BIBER e CONRAD, 2009, p.3), na seara do conhecimento especializado.

⁸ De acordo com a legislação brasileira, a denominação se aplica à categoria de ensino privado sem obrigatoriedade de carga-horária previamente definida; neste caso, faz referência ao ensino de línguas ministrado à parte, em relação ao currículo regular de escolas, faculdades ou universidades.

⁹ *English for Specific Purposes* - Inglês para Propósitos Específicos ou, ainda, Inglês Instrumental, abordagem de ensino motivada pela asserção de que há diferenças significativas entre o uso da língua para propósitos comunicativos gerais e para dar conta de demandas de áreas específicas. O foco pedagógico do ensino de *ESP* é preparar os aprendizes para dar conta dessas diferenças (BIBER, 2006, p. 02). Kennedy & Bolitho (1984) concebem a abordagem a partir de dois pontos principais: o caminho em direção ao foco no aprendiz e a mudança de visão de linguagem, entendida não apenas como um conjunto de regras gramaticais, mas também como um conjunto de funções. Para Dudley-Evans e St. John (1998), *ESP* constitui-se em uma modalidade de ensino centrada igualmente em língua, habilidades, discurso e gêneros apropriados para as necessidades dos aprendizes.

¹⁰ Abordagem de ensino de línguas postulada com base nos estudos funcionalistas de Michael Halliday (1973), que destaca a desenvolvimento da competência comunicativa como objetivo da aprendizagem (RICHARDS *et al.*, 1992, p. 65).

instituição de ensino superior e, a partir disso, pude perceber que a prática da Abordagem Lexical¹¹ ainda era bastante restrita por parte dos materiais instrucionais mesmo nos idos de 2000. Meus primeiros esforços nessa atividade se concentraram na elaboração de unidades didáticas. Permanecia a preocupação com a questão de como desenvolver um trabalho que, da perspectiva pedagógica, pudesse envolver o uso de textos autênticos e levar os aprendizes a reconhecerem sua caracterização e, a partir daí, apropriarem-se do conteúdo específico para seu uso prático.

Por fim, passei a atuar em uma universidade pública voltada para a formação em Ciências da Saúde, em que a leitura de textos científicos e a capacidade de redigir apresentações e *abstracts* de pesquisas acadêmicas em língua inglesa são habilidades cada vez mais importantes, dada a prevalência do inglês na literatura desse campo de conhecimento.

No contexto docente acadêmico, vários colegas professores de áreas aplicadas à saúde costumam solicitar ajuda para a produção de *abstracts* de artigos de pesquisa. Muitas vezes, esses pesquisadores, embora hábeis em produções acadêmicas em língua portuguesa, dão mostras de desconhecer os modos de articulação característicos dos artigos do campo na tradição anglo-saxônica. Ao ministrar uma disciplina de escrita acadêmica em língua inglesa no programa de pós-graduação, ficou evidente a necessidade de, igualmente, auxiliar a comunidade discente participante a promover sua autonomia na redação dos textos. Além disso, no âmbito diretamente dos cursos do campo da saúde brasileiros, o trabalho empírico de exploração do comportamento do léxico a partir de *corpora*, tomando por base as necessidades de aprendizagem específicas dos participantes, ainda apresenta espaço a ser ocupado, para quem se dedica às pesquisas em Linguística Aplicada (LA).

Desse modo, há aspectos lacunares a serem explorados no que tange ao cruzamento entre EAP e Ciências da Saúde, especialmente em termos de pesquisas que enfoquem questões probabilísticas de adoção de padrões lexicais e contexto de uso. Tal empreitada qualifica tanto o ensino de EAP nos cursos de graduação e pós-graduação brasileiros, quanto a formação de educadores linguísticos. É a partir dessa trajetória de trabalho e desse cenário atual de necessidades que esta tese se apresenta.

Pontos de partida e questões de pesquisa

¹¹ Abordagem de ensino de línguas pautada pelo componente lexical, que adota a palavra como unidade de análise e organização de seu *syllabus* (LEWIS, 1993).

Para chegar às questões que a investigação se orienta para responder, partiu-se de alguns pressupostos sobre escrita acadêmica e o funcionamento da língua. Primeiramente, com base na observação previamente mencionada de que colegas pesquisadores fluentes em inglês não eram capazes de reconhecer certos modos de dizer acadêmicos, constatei que fluência não necessariamente garante a escritura de um texto especializado adequado para os parâmetros de uma área de conhecimento, em determinada língua e contexto. Esse ponto me levou a notar que as escolhas do repertório linguístico são, de certo modo, limitadas, a partir de uma perspectiva segundo a qual a linguagem se baseia em padrões organizados pela frequência de uso e não a partir da livre escolha dos usuários¹². Dito de outro modo, não cabe aos pesquisadores adotarem as expressões e formas de dizer que preferirem, porque, além de serem padrão, essas se constituem em blocos. Ao mesmo tempo, para quem desejar ter sucesso em escrita acadêmica em uma língua, é vital reconhecer o que é mais usual adotar para cada contexto. Tais asserções se sintetizam nos seguintes princípios, cujo embasamento teórico será abordado ao longo do texto:

1. A linguagem especializada¹³ funciona em blocos fixos (SINCLAIR, 1991) e se orienta por princípios maiores do que a livre escolha dos usuários;
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 escolha de conteúdo e estilo (SWALES, 1990, p. 58).” Dessa forma, os objetivos comunicativos reconhecidos pelos especialistas influenciam o tema, o estilo 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

¹² 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.

¹³ 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. O tópico *língua/linguagem especializada* será abordado no cap. 2, 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.).

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 e tendo os *abstracts* como objeto de estudo, 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 *lingua 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 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), como os capítulos de revisão de literatura discutirão. Dito de outro modo, a asserção fundamental desta pesquisa é de que aprender a combinar adequadamente as palavras para construir textos em um determinado gênero escrito ajuda o acadêmico do campo da saúde a fazer escolhas linguísticas apropriadas e assim a desenvolver referências acerca do tipo de linguagem que se espera que produza.

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. Além disso, é importante ter acesso a trabalhos descritivos para verificação de aspectos que são constantes e uniformes, bem como para que se possa perceber pontos de variação, cotejando-os uns aos outros, a fim de sistematizar as descobertas, para ensiná-las melhor. Partindo dessas reflexões, este estudo prioriza 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?

Em sequência, especifica-se o objeto da investigação, bem como a unidade de análise adotada. Essa especificação visa *aproximar a lente* das unidades lexicais cuja frequência e caracterização queremos averiguar.

Objetivo de estudo e unidade de análise

O objeto de análise deste trabalho é o texto especializado no ramo das Ciências da Saúde, no macro-gênero *abstract*, mirando os modos de dizer, tanto através dos termos¹⁴ típicos da área e suas conexões, quanto dos aspectos discursivos.

Ao corroborar as investigações de Cabré (2002), assume-se a posição de que está-se tratando de um fenômeno complexo, ao estabelecer o texto especializado como objeto de estudo, já que ele se apresenta maior do que uma única disciplina ou ponto de vista. Ao buscar entendê-lo através de categorizações variadas - tais quais a evolução da Linguística Textual, os estudos de gênero discursivo, o aporte de *registro*¹⁵ e as investigações em Linguística de *Corpus* (LC)¹⁶ - acredito que se tem a chance de percebê-lo como parte de um cenário amplo, de maneira a estabelecer pontos de aproximação entre as perspectivas correlacionadas.

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

¹⁴ 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).

¹⁵ Noção que será desenvolvida no cap. 1 e que abarca escolhas léxico-gramaticais que tipificam as classes de textos, a exemplo dos textos acadêmicos, neste caso.

¹⁶ Área da Linguística que desenvolve estudos empíricos e que, para tanto, se ocupa da coleta e análise de *corpus*. Tal noção será aprofundada no capítulo 4.

tradição dos estudos de Gêneros Acadêmicos, segundo os escritos de Bakhtin¹⁷, Biber e Swales, pelos quais, igualmente, a investigação se orienta, pois esses se unem face a conceberem o texto como um fenômeno com sentido social.

Quanto às ideias de Hoffmann, identifico-me especialmente com o fato de os postulados desse estudioso, acerca dos textos especializados, serem de caráter descritivo, ou seja, ele não pretende submeter o objeto de estudo a regras prévias de como a linguagem *deve se comportar*. Por extensão, alinho-me também ao fato das pesquisas desse investigador promoverem metodologias para o ensino de línguas adicionais.

Para entender a estrutura do texto especializado da forma como nomeado nas questões de pesquisa e chegar aos itens lexicais significativamente mais frequentes, elegi como unidade analítica *lexical bundles* ou *pacotes lexicais*. Esses se referem a “sequências recorrentes de três ou mais palavras independentemente de sua idiomaticidade e de seu *status* estrutural”¹⁸ (1999, p. 990), a exemplo de *'the aim of this'*, *'in the case of'* e *'on the other hand'*, frequentemente encontrados em textos acadêmicos. 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 me identifico (BIBER e CONRAD, 1999; BIBER *et al.*, 1999; BIBER *et al.*, 2004; CORTES, 2004; CONRAD e BIBER, 2005; BIBER, 2006; HYLAND, 2008^c, dentre os principais)¹⁹.

Embasamento teórico

Esta tese pertence ao campo da Linguística Aplicada (LA), uma vez que se estabelece a partir de um posicionamento interdisciplinar, ao buscar promover entendimentos entre a Linguística de *Corpus* (LC), a Linguística das Linguagens Especializadas (LLE) e os Estudos em *English for Academic Purposes* (EAP).

Em sequência, apresenta-se sucintamente um entendimento acerca de como as áreas que cotejadas se relacionam ao objeto de investigação.

A LC é adotada, especialmente sob influência dos estudos em LA desenvolvidos nas universidades britânicas, principalmente em Lancaster, sob influência de

¹⁷ 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, como será discutido no próximo capítulo.

¹⁸ “*Recurring sequences of three or more words recurrent expressions, regardless of their idiomacity, and regardless of their structural status*” (BIBER *et al.*, 1999, p. 990). Todas as traduções constantes neste trabalho são de minha inteira responsabilidade.

¹⁹ Não obstante, existe grande divergência de denominações e definições na literatura sobre combinações de unidades lexicais. O tema será tratado no cap. 4.

Tony McEnery, e, na escola americana, sob a influência de Douglas Biber, na Universidade do Norte do Arizona. Os estudos em LC²⁰ apresentam as seguintes características em comum:

- Um interesse pelo discurso (entendido como a linguagem em ação e em relação a seus usuários;
- Uma ênfase na estreita relação entre vocabulário e gramática;
- Uma preocupação com o uso de dados autênticos e não inventados;
- Uma preferência pelo uso de computadores para análise de grandes arquivos de dados linguísticos.²¹ (GLEDHILL, 2000, p. 10)

Os princípios da LC são, portanto, caros para este trabalho, pois se trata de uma investigação empírica, que se ocupa da coleta e análise de um conjunto de dados linguísticos organizados a partir de *corpora*. Assumiu-se, outrossim, que a orientação para o estudo partiria do *corpus*, ou seja, aceitou-se que seria o *corpus* que apontaria insumos sobre o que deveria ser prioridade ensinar, em se tratando de escrita acadêmica com base nos textos selecionados para compor o acervo.

Partiu-se, portanto, de um acervo de dados que ocupam um lugar no mundo da produção científica, a partir dos quais se organizam considerações sobre a configuração da linguagem. Adotaram-se três sub-áreas de conhecimento como critério de organização, a saber, Medicina, Nutrição e Farmácia, a fim de tornar o *corpus* representativo das produções do grupo prioritariamente beneficiário do estudo, assim como para dar conta das perguntas de pesquisa com mais robustez e promover uma sistematização didática mais ampla. 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 do tema em LC, a exemplo de Gledhill (1995, 2005), Orasan (2001), Hyland e Tsé (2005), Dayrell e Aluísio (2008), Hyland (2004) e Dayrell (2009, 2010, 2011, 2012).

Também foram orientadores os fundamentos da LC, devido ao entendimento de língua como um sistema probabilístico (HALLIDAY, 1992). De acordo com essa noção, é

²⁰ 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. O tema será tratado detidamente no cap. 4.

²¹ “An interest in discourse (*language in action, language in relation to its users*). An emphasis on the close relationship between vocabulary and grammar. A preoccupation with authentic non-invented data. A preference for computers in the analysis of large archives of language.” (GLEDHILL, 2000, p. 10)

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 uma língua não se resume ao preenchimento de espaços vazios de forma incerta, uma vez 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. Em oposição a esse princípio, está o da livre escolha, segundo o qual seria possível selecionar, apenas com restrições gramaticais, os itens que compõem um enunciado. O princípio idiomático mostrou-se inovador e envolveria a seleção simultânea de estruturas maiores, nas quais as palavras não estariam, no caso, sujeitas à discriminação do escritor fluente no domínio de um gênero textual.²² 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 Terminologia²³ centrados nos textos que contêm termos (BOURIGAULT e SLODZIAN, 1999; CIAPUSCIO, 2003; HOFFMANN, 2015), pois este trabalho se ocupa 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. É possível, por exemplo, relacionar certos pacotes lexicais, como ‘*need for further studies*’ com expressões típicas da prosa acadêmica, diferentemente de ‘*need for more studies*’, originária da prosa informal. Esse modo de nomear a ciência, por sua vez, remete à linguagem especializada que Hoffmann (2015) define como “o conjunto de todos os 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). Há uma preocupação com o todo de sentido de um texto, assim como com as escolhas de termos

²² 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.

²³ 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).

típicos através dos quais as comunidades discursivas são identificadas. É em função do 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 (BOURIGAULT e SLODZIAN, 1999; HOFFMANN, 1988; FINATTO, 2004; CONDAMINES, 2005; KRIEGER, 2008; ZILIO, 2012^b) ou ainda como 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. É em função da frequência de manifestações linguísticas específicas que as linguagens especializadas podem ser definidas (HOFFMANN, 2015, p. 40) o que, por sua vez, estabelece um cruzamento com a LC. O foco de atenção em frequência para a LLE abrange o estudo do uso e dos padrões de uso a partir de, entre outros aspectos, combinatórias de palavras mais ou menos fixas, elementos retóricos e argumentativos, colocações especializadas, condições pragmáticas da comunicação e condições do gênero textual. Tais aspectos se relacionam à unidade de análise do estudo.

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 nativos²⁴. Portanto, a LC também é saudada pelos que se interessam por EAP por estar em condições de romper com o paradigma²⁵ tradicional e fornecer aporte ao trabalho de formação de professores-pesquisadores.

Organização do trabalho

Para organizar o recorte proposto e emprestar à pesquisa os contornos desejados, o trabalho se desenvolve em duas partes: a primeira explora os aspectos teóricos do tema e a segunda aborda a aplicação do *corpus*.

Desse modo, a primeira parte do trabalho é composta por quatro capítulos. O capítulo 1 aborda a noção de texto, o substrato da investigação, assim como revê a trajetória dos estudos discursivos e a forma como os textos *in natura* se apresentam. Nesse momento, também introduzem-se os autores através dos quais é proposto um projeto de compreensão teórica para o trabalho. O capítulo 2 se dedica às características através das quais os textos especializados se realizam e se fazem objeto de nosso ingresso ao âmbito do conhecimento científico. Nesse capítulo, trata-se também do campo da LLE por questões de aproximação com os temas analisados. No capítulo 3, apresenta-se a constituição do texto de nosso estudo, o *abstract*, historiando sua representatividade para o mundo acadêmico, caracterizando-o e introduzindo seus principais estudos sob a perspectiva dos textos acadêmicos pela tradição anglo-saxônica. No capítulo 4, abordam-se LC e EAP, dois campos bastante próximos. Nesse capítulo, tratam-se também de pacotes lexicais e discutem-se outros estudos que, igualmente, adotam essa unidade analítica para entender o funcionamento dos textos acadêmicos.

A segunda parte da pesquisa, por sua vez, se organiza em quatro capítulos. No capítulo 5, introduz-se a rota metodológica adotada, apresentando o modo como o *corpus* foi organizado para o estudo, os procedimentos analíticos, a ferramenta adotada para a extração dos pacotes lexicais e o passo a passo do trabalho empírico. Em sequência, no capítulo 6, apresentam-se os dados e as respectivas classificações analíticas. No capítulo 7, desenvolve-se uma síntese dos resultados, a partir da análise do *corpus* de modo comparativo em relação aos acervos maiores que o compõem e respondem-se as questões de pesquisa. Por fim, no capítulo 8, tecem-se considerações didáticas acerca do ensino e do lugar das associações recorrentes

²⁴ 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 (RAJAGOPALAN, 2004; WILLIAMS, 2006).

²⁵ Adota-se o termo *paradigma* aqui para referir a uma aceção prescritiva.

de palavras em face à produção de *abstracts* e produzem-se considerações a respeito das perspectivas e limites da investigação .

I. PRIMEIRA PARTE - A FUNDAMENTAÇÃO TEÓRICA

1. CAPÍTULO 1

“Portanto, por trás de cada texto está o sistema da linguagem. A esse sistema correspondem no texto tudo o que é repetido e reproduzido e tudo que pode ser repetido e reproduzido, tudo o que pode ser dado fora de tal texto (o dado). Concomitantemente, porém, cada texto (como enunciado) é algo individual, único e singular, e nisso reside todo o seu sentido (a sua intenção em prol da qual ele foi criado). É aquilo que nele tem relação com a verdade, com a bondade, com a beleza, com a história.”

BAKHTIN, 2003, p. 310

TEXTO, O TECIDO DA INTERAÇÃO

Para iniciar a exploração dos elementos a partir dos quais a pesquisa se organiza, é importante dizer que toda investigação parte de pontos de vista que sintetizam a nossa história como pesquisadores, nossas crenças e desejos. Quando se recorta um tema de estudo, o contorno assumido em função de escolhas feitas (de)limita o objeto pesquisado, e ele vai adquirindo facetas com as quais nos identificamos.

Assim, dentre as várias perspectivas que estudiosos de questões linguísticas vêm adotando para se aproximarem de seu objeto de análise, parte-se aqui do entendimento de língua na condição de atividade social, histórica e cognitiva (Bakhtin, 1997)²⁶. É com a preocupação de conceber a língua como forma de aproximação entre as pessoas em um mundo cada vez mais especializado que esta investigação se volta para o estudo dos textos acadêmicos.

Desse modo, para aqueles de nós com familiaridade no âmbito das práticas sociais em leitura ou mesmo em escrita acadêmicas - geralmente pesquisadores, professores ou estudantes - que partilhamos saberes próprios de nossas áreas de conhecimento, o contato com um texto especializado comporta expectativas quanto ao formato ou mesmo quanto ao modo como o conteúdo costuma ser desenvolvido. Tais características diferenciam a produção científica especializada, objeto de estudo deste trabalho, de um texto literário, de uma crônica policial ou de uma liturgia religiosa, por exemplo. De acordo com Giering, Alves e Mello (2010), que adotam uma analogia com o contrato social²⁷ para explicar as distinções entre as classes de texto, “os contratos diferem porque são diferentes as identidades dos que se comunicam, os objetivos que querem alcançar, os temas, os meios” (p.7).

Ao longo de nossa interação com o mundo escrito especializado, vamos construindo competência para o reconhecimento das convenções, regras e normas que regem os ‘contratos’ da ciência. A experiência na identificação das marcas típicas dos registros do conhecimento humano, por sua vez, favorece o acesso à interpretação e à produção dos discursos veiculados socialmente. Estudiosos do campo das práticas textuais denominam tal experiência de letramento²⁸ (ROJO, 2012; SOARES, 2003) ou mesmo de leitura de mundo (FREIRE, 1989). Quanto mais ampla seja a condição de letramento de um indivíduo, tanto

²⁶ Entendimento compartilhado por Douglas Biber (1988) e John Swales (1990).

²⁷ “Documento através do qual se estabelecem direitos e obrigações recíprocos aos seus signatários” (DAMIN, ALVES, BECKER, SOUZA, GIERING, ALBÉ, KISSMANN & MELLO, 2013, p. 7).

²⁸ Nos limites deste trabalho, entende-se letramento como “práticas sociais de leitura e de escrita mais avançadas e complexas que as práticas do ler e do escrever resultantes da aprendizagem do sistema de escrita” (SOARES, 2004, p. 6).

mais vasta é sua possibilidade de exercer sua cidadania e atuar com plenitude no mundo em que vive.

Neste capítulo, iniciamos a análise do objeto de estudo da pesquisa, abordando a noção de texto, elemento básico da comunicação humana, através de uma revisão da trajetória dos estudos textuais. Discute-se também o modo como os textos *in natura*, isto é, textos autênticos, se constituem em gêneros para dar forma à interação. Por fim, apresentam-se os autores através dos quais se propõe um projeto de compreensão teórica para a pesquisa, uma vez que os postulados desses investigadores sintetizam o modo como o trabalho se posiciona face aos estudos discursivos.

1.1. Texto, *textum*, tecido

Originária do latim, a palavra *texto* remonta a *tecido*, *entrelaçamento*, como nos ensinou Roland Barthes (1973/1996), pesquisador que reivindicou, para cada texto e para cada leitura, a condição de serem únicos em suas diferenças.

Muito embora os estudos do tema, pela perspectiva da Linguística, sejam relativamente recentes, o interesse por desvendar essa trama é uma herança grega. Três séculos antes do começo da nossa era, Aristóteles já se dedicava aos textos orais e escritos, especialmente à retórica²⁹, embora os estudos da época fossem desenvolvidos a partir de um caráter formal ou essencialista.

Em uma perspectiva diacrônica, foi apenas a partir do século XX que autores como Bakhtin (1986; 1997), Bazerman (1988) e Swales (1990) passaram a defender a ideia de que aquilo que produzimos em nossa comunicação, seja ela oral ou escrita, não escapa a algum tipo de gênero. Por conseguinte, nossa comunicação se encontra atrelada a alguma classe de texto, como uma espécie de teia, que vamos tramando para nos comunicar e à qual não podemos escapar.

Recurso central não apenas para a representação do conhecimento, mas também para a sua construção individual e coletiva, o texto já passou por várias definições, de acordo com a abordagem adotada. Nesta investigação, amparada na Linguística de *Corpus*,

²⁹ Entendida, neste contexto, como “o estudo de como as pessoas usam a linguagem e outros símbolos para realizar as ações humanas” (BAZERMAN, 1988, p.6). “*The study of how people use language and other symbols to realize human goals and carry out human activities.*”

toma-se como referência a concepção de texto como linguagem em uso (HALLIDAY e HASAN, 1978), em função da dimensão social e interativa compreendida nesse conceito.

Segundo Brinker (2005), há duas correntes principais para o estudo do texto: abordagens que focalizam a análise na língua como sistema, que se relacionam a um interesse pelas normas de funcionamento, e abordagens que analisam o texto sob uma perspectiva comunicativa, ou seja, no que tange à relação com o mundo do interlocutor ou com o contexto do evento comunicativo. Nesta pesquisa, busca-se correlacionar esses dois aspectos.

Produto da atividade verbal humana, o texto se torna unidade básica dos estudos linguísticos de forma heterogênea, em face das várias orientações investigativas, às quais correspondem propostas teórico-metodológicas diversas. Nosso estudo, por sua vez, tem pontos de contato com a linha da Linguística ou Pragmática Textual Aplicada³⁰, cujas investigações se dedicam à interação empírica dos indivíduos com textos cotidianos (BLUHDORN e ANDRADE, 2005), como é interesse explorar aqui.

A Linguística Textual ou Teoria do Texto constitui uma subdisciplina da Linguística de tradição predominantemente europeia. Esse ramo da Linguística é originário dos anos 60, do século XX, principalmente com base na Alemanha, Países-Baixos, na França, no Reino Unido e nos Estados Unidos. O objeto específico da Linguística Textual são os processos de construção de textos, assim como os fatores envolvidos em sua produção e recepção, por meio dos quais os participantes do ato comunicativo criam sentidos e interagem com outros seres humanos. Dada a amplitude dos estudos do campo, na avaliação de Koch (1997), uma das principais representantes brasileiras do ramo da Linguística Textual, a passagem por diferentes momentos de inspiração em modelos teóricos variados, como mencionamos acima, “não deixa de ser bastante natural em uma ciência em formação” (p. 67).

Nos anos 80, a Linguística Textual recebeu considerável receptividade em vários países do mundo, dentre eles, o Brasil³¹ (BLÜHDORN & ANDRADE, 2005). Por conta dessa vertente, o esforço em entender o texto gerou, em linhas gerais, três níveis de complexidade para os estudos do campo: gramatical, semântico e pragmático³².

O primeiro nível de complexidade corresponde a uma fase denominada *transfrástica*, assim denominada por referência às análises interfrasais características dessa etapa. O interesse do período era estudar aspectos como a pronominalização, a seleção de

³⁰ A Linguística Textual Teórica se volta para o estatuto do texto (BLUHDORN e ANDRADE, 2005).

³¹ Vide trabalhos de Fávero e Koch (1983), Koch (2004), Marcuschi (1983) e Neiss (1981).

³² Embora, alguns estudiosos, a exemplo de Conte (1977), defendam que essa distinção seja mais da ordem tipológica, alinhando-me a Koch (1997), uma vez que essa pesquisadora reflete que “há, sim, uma cronologia envolvida nesta sucessão” (p. 68).

artigos e os tempos verbais, dentre outras características gramaticais. O percurso ainda era da frase para o texto, com base na constatação de pesquisadores dessa vertente de que existem fenômenos gramaticais que não podem ser explicados no nível da oração apenas, pois esta não dá conta de esclarecê-los. Essa unidade de análise foi o ponto de partida para pioneiros do campo, como Harweg (1979 [1968]), que definiu o texto como uma sequência de unidades linguísticas mediante encadeamento pronominal ininterrupto. Representantes dessa corrente postularam que o texto, e não a oração, representa a unidade básica da estrutura da língua (HARTMANN, 1971; WEINRICH, 1993).

Os esforços do período para desenvolver uma *linguística da frase ampliada* acabam sendo abandonados, no entanto, de acordo com Koch (*op. cit.*) por haver entre frase e texto uma diferença de ordem qualitativa e não quantitativa. Abre-se, então, espaço para a constituição da etapa seguinte.

A perspectiva semântica inauguraria a segunda fase dos estudos em Linguística Textual, com forte influência do Gerativismo, corrente que viria a se tornar muito difundida (CHOMSKY, 1986). O Gerativismo se volta para a competência linguística do falante, a partir da noção de que este tem a propriedade inata da linguagem e é capaz de produzir enunciados infinitos e inéditos em uma dada língua. De forma análoga, essa fase passou a levar em consideração a competência textual do falante. A constituição textual era compreendida por um sistema de regras, baseado em regularidades, assim como o texto passa a ser compreendido como o “signo linguístico primário” (HOFFMANN, 2015, p. 47), a partir do qual se chega a unidades menores.

Desse entendimento do texto como instância dotada de unidade própria se originaria a noção de *gramáticas textuais*, que têm por meta explicar o que a gramática sentencial da fase transfrástica não conseguia dar conta. Marcuschi (1998) analisa que as gramáticas textuais propuseram o texto como o objeto central da Linguística e assim procuraram estabelecer um sistema de regras finito e recorrente, partilhado por todos os usuários de uma língua. Esse sistema de regras habilitaria os usuários a identificarem se uma dada sequência de frases constitui ou não um texto e se esse texto é bem formado. Observe-se que essa fase adota um método de análise descendente: do texto rumo às unidades menores que o constituem, a fim de categorizá-las. No entanto, a ideia era de que se mantivesse intacta a função textual dos elementos individuais, uma vez que o texto não poderia ser reduzido a “uma sequência de cadeias significativas” (KOCH, 1997, p. 69).

Além dos gerativistas, nesse período, destacam-se os autores estruturalistas, dentre eles Harald Weinrich, que define o texto como um “andaime de determinações” em

que tudo está necessariamente interligado (1993). Para Weinrich, toda Linguística representa Linguística do Texto. Distingue-se também o trabalho do holandês Van Dijk, que vem sendo referido como um dos fundadores da disciplina. Nos anos setenta, Van Dijk abandona o projeto das gramáticas textuais para se dedicar às macroestruturas textuais e à descrição de novas classes de texto, dentre as quais, o relato científico (VAN DIJK, 1979, 1980). Esse estudioso igualmente se destaca por voltar-se para a noção de contexto e estabelecer uma distinção entre discurso e texto, percebendo o primeiro como texto contextualizado.

Nessas duas primeiras fases descritas, o texto era então percebido como um produto acabado, enfatizando-se a materialidade linguística, o aspecto formal do discurso. Os estudos eram teóricos e estavam mais voltados ao estatuto sistemático do texto propriamente dito, funcionando um pouco como um mapeamento do que era entendido como território textual. Essas investigações tiveram por base microtextos canônicos, prototípicos³³, muitas vezes partindo de idealizações, não raro criados para as próprias pesquisas e em uma só língua, como os de Harweg (1979) e dos funcionalistas Halliday & Hasan (1976). Cumpre notar que esses dois últimos autores contribuíram para o campo com o conceito basilar de coesão³⁴ textual, que se tornaria uma espécie de emblema desse período dos estudos textuais.

Outrossim, a própria noção de gramáticas textuais, características da fase semântica, partia da separação da unidade estrutural gerada a partir da competência de um usuário idealizado e descontextualizado, de modo que o texto não era ainda compreendido a partir do uso em uma situação real de interação.

Apesar de limitações que podem ser constatadas com base nas considerações tecidas acima, houve pioneirismo nessas fases, e a Linguística Textual florescia como campo de estudos. Não obstante, a evolução do campo apontava para a elaboração de uma teoria de texto que discutisse a constituição, o funcionamento e a produção dos textos em uso em situação real de interação verbal e que viria a se consolidar na fase seguinte.

O próximo período inauguraria a Linguística Textual Aplicada propriamente dita. A partir dessa fase, os textos começam a ser explorados com base em seu contexto, o que abrange também o conjunto de condições externas, tais quais os fatores que intervêm em sua

³³ A noção de prototipicidade é tomada emprestada de teorias semânticas, a pioneira e mais conhecida delas sendo a de Rosch (1978).

³⁴ A noção de coesão trata basicamente das articulações gramaticais existentes entre as palavras, as orações e frases para garantir uma boa sequenciação de eventos e uma harmonia entre os elementos de um texto. Halliday & Hasan (*op. cit.*) se debruçaram sobre as relações de sentido existentes no interior de um texto, as quais se materializam através de determinados recursos léxico-gramaticais, tais quais, por exemplo, conectores, os campos lexicais e as colocações, mecanismos que dão tessitura ao texto. É com base na confluência desses elementos que esses dois autores traçaram seu conceito de coesão.

produção, recepção e interpretação. Essa fase se denomina de pragmática.

A fase pragmática, contribuiu com os estudos acerca dos padrões de textualidade de autoria de Beaugrande e Dressler (1981), que concebem o texto como um documento de decisões, de processos de escolha e combinação, uma ocorrência comunicativa. Com base principalmente na noção de coesão textual, enfocada sob a perspectiva de Halliday e Hasan (1976) já mencionada, Beaugrande e Dressler apontaram sete padrões ou critérios de textualidade que tornam os textos comunicativos. Dentre esses, os mais enfatizados nessa década foram os dois primeiros: coesão e coerência³⁵, que estão centrados no próprio texto por designarem operações direcionadas para os materiais linguísticos. Os demais padrões se voltam para o usuário, são eles informatividade, situacionalidade, intertextualidade, intencionalidade e aceitabilidade. Todos esses aspectos trabalham para distinguir um texto de uma sequência de palavras, pois recorrem para que um texto exerça suas “funções comunicativas e interacionais, complementando os princípios da coesão formal e da coerência semântica” (BLÜHDORN e ANDRADE, 2005, p. 18). Beaugrande inclusive se referiria ao texto como acontecimento discursivo para o qual convergem ações linguísticas, sociais e cognitivas (1997).

Nessa fase, em lugar de procurar descrever a competência textual do falante, como na anterior, tornou-se mais viável analisar de que forma se constituem e funcionam os textos em uso, bem como o modo pelo qual se dá a sua compreensão. Assim, surgiu o período de elaboração de uma teoria do texto. Passou-se a perceber o texto não mais como um produto acabado e sim como um processo que resulta de questões sócio-cognitivas, interacionais e comunicativas. Em outras palavras, o texto passa a ser entendido como um ato de comunicação complexo.

A partir dessa terceira ênfase, os estudos textuais têm se deslocado da teoria para a aplicação. A possibilidade de trabalhar com grandes coleções de dados computadorizados, a partir do desenvolvimento da Linguística de *Corpus* (LC), permitiu a aproximação de textos não canônicos, de extensões variadas, autênticos e não idealizados. Tais textos podem inclusive ser incompletos ou apresentar imprecisões, no que diz respeito a aspectos internos, como é possível verificar nos dados que serão apresentados para análise, por exemplo. Em que pese saber, uma particularidade da análise textual pela perspectiva da LC é de que o texto, em alguns momentos, transforma-se em *corpus*, o que implica dizer que pode perder as características do todo para se submeter a microanálises de suas características

³⁵A noção de coerência trata da consonância entre referentes textuais e situacionais ou a ausência de contradições, que faz com que o que é dito em um texto faça sentido como um todo.

(FINATTO, 2011^a).

Na fase pragmática, passou-se a explorar de forma conjunta fatores gramaticais, semânticos e pragmáticos, destacando-se por vezes as características formais dos dados, por vezes os traços conceituais-cognitivos ou os atributos comunicativos e interacionais, de acordo com o enfoque empírico e com as interações práticas das pesquisas. Representante dessa visão eclética do texto, Ciapuscio (2003), linguista argentina afiliada à corrente dos estudos em Terminologia de Perspectiva Textual³⁶, designa o termo *texto* como “produtos verbais orais e escritos, em toda sua complexidade, isto é, incluindo, além da dimensão estritamente linguística, as dimensões funcional-comunicativas” (p. 38). Essa autora postula que, para compreender um texto, colocamos em jogo vários sistemas de conhecimentos inter-relacionados: nosso conhecimento de mundo ou enciclopédico, ou seja, o que conhecemos sobre o tema ou área do texto, assim como o conhecimento linguístico, aspectos formais da língua, que envolvem léxico e gramática e ainda o nosso conhecimento interacional-situacional, em uma alusão a como nos situamos acerca do tema em pauta.

A partir de uma ênfase nesses sistemas inter-relacionados, de acordo com trabalho clássico de Koch (1995), os participantes coordenam suas ações com base nas “condições sob as quais a atividade verbal se realiza” (p. 22) e, no curso da interação, constroem sentido para os textos que produzem, a partir do conhecimento que é compartilhado. A produção textual, por sua vez, constitui-se em uma atividade interativa verbal com fim social, orientada para os parceiros da comunicação. Está implícita aqui a visão de que o sentido não está no texto, mas é construído ao longo da interação, ou seja, somos nós, leitores, que o desenvolvemos ao entrarmos em interação com os textos que lemos³⁷. Compartilhando o entendimento de Ciapuscio (2003) mencionado, Koch (2002) também reforça o ponto de que, para extrairmos sentido de um texto, “faz-se necessário recursos aos vários sistemas de conhecimento, assim como a ativação de processos e estratégias cognitivas e interacionais” (p. 26).

De forma resumida, através dessa breve história das fases dos estudos textuais, podemos dizer que dar conta *desse todo que faz sentido, que simboliza algo e que comunica* (BARROS, 1990) implica lidar com uma complexidade de camadas sobrepostas de elementos linguísticos e comunicativos, que se organizam para dar forma à comunicação. Agora, cumpre

³⁶Vertente controversa do campo da Terminologia que ainda está dando os primeiros passos no Brasil. Os pressupostos dos estudos de Terminologia de Perspectiva Textual serão discutidos no cap. 2.

³⁷Para aprofundamento dessa noção, recomenda-se a leitura da teoria de Fish (1976) acerca comunidades interpretantes. Esse autor considera que o texto não é autossuficiente e que, portanto, é a experiência do leitor, ao invés das estruturas do texto, que deveria ser objeto de investigação.

discutir como essas camadas se organizam para dar forma aos textos.

1.2. Classes de eventos relativamente estáveis e situacionais

Primeiramente, é importante destacar que o conhecimento de mundo dos leitores, que inclui o reconhecimento das formas e funções dos textos, é uma das principais fontes de coerência textual referida acima. Isso porque os textos não *funcionam* de modo isolado; suas formas e funções podem ser comparadas a outros textos que desempenham papéis semelhantes e que apresentam características estruturais equivalentes (HEUBOECK, 2009). Os textos, desse modo, também representam modelos de instâncias comunicativas. “A teoria dos gêneros reconhece essas classes de textos como modelos comunicativos”³⁸. (p. 36). Entendo, na verdade, que a teoria dos gêneros vai além. É com base na noção de gênero que esta parte de nosso texto se apoia.³⁹

Seguindo a trajetória sobre estudos textuais rumo ao mundo da escrita especializada, é centralmente a partir dos postulados que tiveram origem em três autores que esta parte de nossa análise se organiza: Mikhail Bakhtin, John Swales e Douglas Biber.

A escolha por Bakhtin, filósofo e pensador russo que se debruçou sobre o tema da linguagem, se dá porque as ideias desse linguista russo, que é um dos primeiros e mais influentes estudiosos de gêneros do discurso, têm sido tomadas como ponto de partida para grande parte das investigações no tema. A adoção dos pressupostos de John Swales ocorre porque esse autor inglês se volta para os gêneros no contexto aplicado a ensino/aprendizagem de inglês para propósitos acadêmicos. As contribuições de Douglas Biber, por sua vez, devem-se ao fato de que esse linguista norte-americano, que também se dedica ao estudo dos gêneros acadêmicos, alargou essa noção em suas pesquisas, agregando a perspectiva de *registro* (BIBER, 1995), a qual é também importante para nosso estudo.

A motivação, portanto, parte de uma identificação com a concepção que esses autores compartilham de que nada na comunicação humana, de fato, escapa a alguma classe de gênero. Essa noção generificada para a interação é muito mais ampla do que uma forma estrutural ou um modelo, pois se trata de uma maneira de conceber a linguagem e a

³⁸ “Genre theory refers to such classes of texts as communicative genres.” (HEUBOECK, 2009, p. 36).

³⁹ Afora o Círculo Bakhtiano (década de 1920), três escolas ou tradições de análise de gêneros textuais são reconhecidas no ocidente: New Rhetoric (BAZERMAN, 1988), Registro e Teoria de Gênero (influenciada pela Linguística Funcional Sistêmica de Halliday) e a abordagem de ESP/EAP (Swales, 1990; Bhatia, 1993; Hyland, 2002). Não é, contudo, propósito deste estudo ater-se às diferenças entre essas escolas.

organização dos textos orientados para possibilitar uma infinidade de modos de aproximação entre as pessoas. Ao mesmo tempo, tal concepção se completa diante do fato de os três investigadores nomeados compartilharem de um entendimento de língua como sendo um empreendimento social, histórico e cognitivo (BAKHTIN, 1997), como já havia sido registrado anteriormente. Desse modo, o pensamento desses estudiosos é uma síntese dessas representações, na forma como se apropriam da linguística do texto e de toda uma história do pensamento linguístico. A proposta, ao conjugar suas perspectivas, portanto, é de apresentar um projeto de compreensão teórica para este trabalho.

Antes de abordarmos a síntese de como esses três autores entendem as formas de realização social dos textos, um aspecto conceitual e terminológico importante é entender a diferença entre as noções de gênero, tipo e classe textual.

Sucintamente, o termo *gênero textual*, com base em uma interpretação da noção seminal de Bakhtin (1999), refere-se a inúmeros textos que somos capazes de reconhecer porque se repetem de modo razoavelmente fixo nos contextos sociais de nossa identificação, a exemplo de *abstracts*⁴⁰, teses, dissertações e resenhas acadêmicas, dentre tantos outros. Linguistas alemães chegaram a nomear mais de quatro mil gêneros. Acredita-se que a evolução tecnológica, por sua conta, contribua para a criação de mais e mais novos gêneros (MARCUSCHI, 2006).

Já *tipo textual*, de acordo com Biber (1988), Swales (1990) e Marcuschi (2006), serve para designar uma classificação científica que se relaciona a algumas poucas categorias como narração, argumentação, exposição, descrição, injunção, ou seja, “uma espécie de sequência teoricamente definida pela natureza linguística de sua composição (aspectos lexicais, sintáticos, tempos verbais, relações lógicas)” (MARCUSCHI, 2006, p. 27).

A denominação *classe textual*, por outro lado, remonta ao que sabemos enquanto falantes de uma língua; “classificações cotidianas que podem ser mencionadas por meio de determinados lexemas condensadores do saber sobre determinada classe textual” (CIAPUSCIO, 1994, p. 25), a exemplo de se classificar ‘isto é uma crônica’, ‘isto é uma receita gastronômica’.

1.2.1. Das contribuições de Mikhail Bakhtin

⁴⁰ É importante dizer, no entanto, que Swales e Feak (2009) estabeleceram distinção entre gênero e sub-gênero ou gênero parcial do discurso e que atribuem o *abstract* a essa última categoria. Outrossim, defende-se aqui que o *abstract* se sustenta como gênero textual, aspecto que será retomado no cap. 3.

Do mesmo modo como os textos são elementos básicos da comunicação humana, para Bakhtin (1986; 1997), é através de algum tipo de gênero que tais elementos tomam forma. De Bakhtin (1953:1997), partiu o entendimento de gêneros como entidades sócio-discursivas e formas de ação social adotadas para agir sobre o mundo e dizer o mundo. Para ele, os gêneros são elaborados em diversas esferas da atividade humana e são capazes de refletir as condições específicas e as finalidades de cada uma dessas esferas. Isso porque, até então, eles haviam sido estudados apenas pela perspectiva da retórica e da literatura, em detrimento de estarem presentes em toda forma de interação verbal. As ideias desse pensador, para quem linguagem e sociedade são duas esferas indissociáveis, representaram uma renovação no pensamento linguístico ocidental, ao cruzarem as fronteiras na década de setenta.

Bakhtin (1979:1997) nomeou os gêneros do discurso como “tipos relativamente estáveis de *enunciados*⁴¹”, elaborados de acordo com as condições específicas de cada campo da comunicação verbal. Em suas palavras,

Todos os diversos campos da atividade humana estão ligados ao uso da linguagem. Compreende-se perfeitamente que o caráter e as formas desse uso sejam tão multiformes quanto os campos da atividade humana, o que, é claro, não contradiz a unidade nacional de uma língua. O emprego da língua efetua-se em forma de enunciados (orais e escritos) concretos e únicos, proferidos pelos integrantes desse ou daquele campo da atividade humana. (BAKHTIN, 2003, p. 261.).

Note-se que, para ele, há uma forte correlação entre linguagem e conexão do homem com o meio, o que parece relacionar o uso aos modos de transformar o mundo ou, pela perspectiva de nosso interesse, de especializá-lo. Além disso, ele contextualiza a multiplicidade de usos da linguagem também em função da atividade humana. Esse entendimento é muito próximo ao dos estudiosos da linguagem especializada, para condicionar os fatores temáticos e pragmáticos que determinam a realização linguística.

Ainda, através das palavras de Bakhtin (*op. cit.*) citadas acima, temos que a unidade de comunicação entre os participantes é o enunciado. Esse é composto por diversas

⁴¹ A noção de *discurso* concebida pelo Círculo Bakhtiniano, Escola de linguistas russos fundada por Bakhtin, originalmente abarca a essência da língua; a discursividade, considerando primordialmente a construção dos sentidos. O discurso se realiza na forma de enunciados, que poderiam equivaler, *grosso modo*, às palavras ou orações produzidas de modo autoral, porque proferidas de dentro de um discurso. À exemplo da palavra *lindo*, que pertence à classe gramatical dos adjetivos e que, quando proferida diante de um quadro de que gostamos (*Lindo!*), ganha autoria e endereçamento e, portanto, se torna um enunciado (MARCUSCHI, 1996; BAKHTIN, 1997).

vozes, que dialogam dentro do discurso, uma vez que o que dizemos, muitas vezes, é uma reorganização do que alguém já disse ou terá eco em outras vozes, como se fôssemos ventríloquos sociais (1986). Esse *eco social* ou polifonia, significa dizer que uma característica do funcionamento discursivo é a presença de várias instâncias enunciantes ou de vários textos dentro de um mesmo texto. No caso dos artigos acadêmicos, essa concepção pode ser identificada, por exemplo, pelo uso de citações ou referências a outros autores ou obras. O centro organizador da enunciação, por sua vez, “não é interior mas exterior: está situado no meio social que envolve o indivíduo” (BAKHTIN, 1999, p.118).

Muitas vezes, contudo, a noção de enunciado aqui se equivale à noção mais comum de texto. Vários pesquisadores do campo dos estudos dos gêneros, porém, como Hasan (1978) e Marcuschi (2006), empregam o termo *texto*, em vez de enunciado, ao se referirem à noção de gênero dentro da mesma perspectiva social de uso da língua que Bakhtin⁴².

Os gêneros resguardam determinadas características pelas quais são identificáveis, tais como a forma de distribuição de conteúdo na página, a estruturação composicional, o estilo, as escolhas linguísticas e o público a quem se destinam. Por serem eles estabelecidos socialmente; isto é, reconhecidos e definidos por culturas e comunidades específicas, costumam ser culturalmente sensíveis em relação aos critérios de prototipicidade⁴³. Além disso, também cumprem determinada função social nas relações humanas, porque estabelecem expectativas nos interlocutores, a exemplo do *abstract*, cuja leitura se dá em função de expectativas dos leitores de encontrarem determinadas informações acerca da pesquisa a que o texto se refere⁴⁴. Podem ter também suporte na forma ou mesmo no ambiente em que um texto apareça. No caso do artigo acadêmico, sua publicação em uma revista especializada também é o pano de fundo de sua caracterização. No entanto, não se pode tampouco supervalorizar a forma de um texto para determinar-lhe o gênero. Uma entrevista, ao ser registrada em um artigo acadêmico, por exemplo, torna-se uma parte do gênero maior que a engloba.

Bakhtin acrescenta que os gêneros nos são dados como nos é dada nossa língua materna, ou seja, nós a adquirimos mediante enunciados concretos que ouvimos e reproduzimos durante a comunicação verbal que se efetua com os indivíduos que nos rodeiam (1952-1953/1979). Eles se complexificam e tornam-se instrumentos de construções novas,

⁴² Neste trabalho, transpõem-se *gêneros do discurso* para *gêneros textuais* sem maiores distinções.

⁴³ Um exemplo disso, como será abordado no cap. 3, é o fato de diferentes publicações especializadas, e mesmo campos do saber, definirem critérios distintos para produção de *abstracts*.

⁴⁴ Note-se que essa noção remonta à teoria de Fish, aludida na seção anterior.

mais elaboradas. Os gêneros primários remetem a palavras, frases e expressões, são aqueles constituídos da comunicação verbal cotidiana, enquanto os secundários aparecem em circunstâncias de comunicação mais complexas, principalmente escritas, as que nos interessam aqui.

Compreende-se que, durante o processo de formação de gêneros secundários, denominados de *transmutação*, esses incorporam e transformam os gêneros primários. Ao serem absorvidos, esses últimos perdem sua relação imediata com a realidade existente com os discursos alheios, conservando sua forma e significado cotidiano apenas no conteúdo. A integração com a realidade ocorre, então, através do gênero que o incorporou. Observe-se que essas novas construções não resultam direta e necessariamente da esfera de motivações do participante, da esfera de suas experiências pessoais, mas de um mundo outro que tem motivações mais complexas por construir. Tais motivações não são mais necessariamente pessoais e suas formas de materialização precisam ser aprendidas, para que possamos tomar parte desse mundo; em nosso caso, o mundo das linguagens especializadas.

A diferença entre esses dois gêneros básicos não deve ser ignorada porque

ignorar a natureza dos enunciados e as particularidades do gênero que assinalam à variedade do discurso em qualquer área do estudo linguístico leva ao formalismo e à abstração, desvirtua a historicidade do estudo, enfraquece o vínculo existente entre a língua e a vida. A língua penetra na vida através dos enunciados concretos que a realizam, e é também através dos enunciados concretos que a vida penetra na língua. (BAKHTIN, 1992, p. 283)

Um dos aspectos mais interessantes e originais desse postulado parece ser o fato de que ele valoriza os enunciados concretos, em vez de tomá-los como singelos ou menores, porque são concretos. Ao contrário, ele confere valor à trajetória do indivíduo e à variedade do discurso que forma nossa história textual e que faz de nós seres capazes de transitar pelos diferentes gêneros.

Apesar do fator de previsibilidade e da capacidade interpretativa das ações humanas, os gêneros, não são estanques ou enrijecedores da criatividade textual. Alguns deles apresentam inclusive ancoragem em outros gêneros (a exemplo do *abstract* em relação ao texto científico⁴⁵ que lhe dá origem), como notado por Bakhtin (1997). Tal caracterização, também acaba por delimitar os gêneros como fenômenos sócio-históricos, relacionados de perto à vida cultural e social.

⁴⁵ A propósito, Finatto (2011a) nota que “o texto científico só se coloca como tal, institucionalmente, a partir dos anos 1930, quando ocorre uma primeira reunião internacional dos editores de textos científicos” (p. 32).

Todos esses aspectos funcionam a serviço de atestar que os gêneros são eventos dinâmicos com definições nem sempre unívocas. Muito mais do que formas linguísticas, nas concepções de Bakhtin, temos que os gêneros são modelos comunicativos e, portanto, formas de realizar linguisticamente ações sociais.

1.2.2. Das contribuições de John B. Swales

Em uma perspectiva de compreensão do fenômeno dos gêneros semelhante à estabelecida por Bakhtin, porque voltada para a utilização social da língua, estão as contribuições teórico-metodológicas de John Swales. Esse linguista explicou gêneros como classes de eventos comunicativos compartilhados pelos membros de uma determinada comunidade.

Observemos que a definição de Swales parece reforçar o papel do *outro* introduzido por Bakhtin, na figura do membro com quem se compartilha o texto, aquele que está na mesma esfera comunicativa que nós. Cumpre dizer, no entanto, que esse autor não se baseou nas concepções bakhtinianas para desenvolver seus postulados, embora já tenha declarado sentir-se “completa e dolorosamente culpado” (SWALES, 1992) por não ter dado mais atenção à Bakhtin, cujo trabalho conhecia apenas superficialmente.

Swales teceu sua definição de gênero, em princípio, voltando-se pontualmente para o ensino e a aprendizagem da língua inglesa para finalidades acadêmicas. Ele o fez nos seguintes termos:

Um gênero envolve uma classe de eventos comunicativos, dos quais os membros compartilham um conjunto de objetivos. Esses objetivos são reconhecidos pelos membros especialistas da comunidade discursiva de origem e, dessa forma, constituem a razão para o gênero. Essa razão formata a estrutura esquemática do discurso e influencia e delimita a escolha de conteúdo e estilo (SWALES, 1990, p. 58).

Com base nessa definição, tem-se que, tanto para Swales, quanto para Bakhtin, os gêneros são concebidos a partir de objetivos comunicativos que influenciam o tema, o estilo e a estrutura do texto. A diferença parece ser pela inclusão dos membros peritos por

Swales, que não se encontra na abordagem de Bakhtin⁴⁶, para quem é possível que esse reconhecimento de quem legitima um gênero ocorra implicitamente pelos atores sociais, com base nas necessidades comunicativas.

Professor de escrita acadêmica para aprendizes da modalidade *ESP/EAP*, Swales buscou ancoragem (a) nas abordagens voltadas para o ensino funcional da língua inglesa; (b) nas quatro habilidades da língua, dentre essas, especificamente, nas estratégias de leitura; (c) na abordagem nocional-funcional⁴⁷ bem como, (d) na análise do discurso anglo-saxônica e crítica e, por fim, (e) nas concepções do antropólogo Clifford Geertz⁴⁸, no que tange a formas de conhecimento do mundo quanto ao contexto e à produção de conhecimento.

A análise linguística empregada por Swales mira a construção do texto e as práticas sociais que norteiam as escolhas linguísticas que lhe dão forma. Nessa perspectiva aplicada, Swales preconiza a importância dos gêneros de textos acadêmicos serem amplamente descritos, a fim de que se possa desenvolver materiais de aprendizagem e metodologias adequadas ao ensino.

Em função dos gêneros acadêmicos apresentarem grande variedade, Swales (2004) refere-se a eles como constelações, em uma alusão ao aspecto de que quase nunca os encontramos de forma isolada. Os *abstracts*, por exemplo, em boa medida, associam-se aos trabalhos científicos a que se referem.

Para a mobilização de conhecimentos com vistas à produção e/ou compreensão textual, o trabalho com gêneros é fundamental para o aprendiz. Afinal desenvolver fluência em práticas de escrita e compreensão, por exemplo, implica refletir sobre uma série de características comuns aos textos de um dado gênero. Os estudos desse pesquisador têm dado suporte a aprendizes dos textos acadêmicos na tradição anglo-saxônica para identificarem “características formais e funcionais e para que desenvolvam a sua capacidade de produzir

⁴⁶ Uma vez que Bakhtin não se debruçou sobre o texto científico. No entanto, que Bazerman, por sua vez, parece estar subsumido nessa concepção, embora, como será discutido um pouco adiante, não tenha partido dos mesmos pressupostos teóricos e não tenha a intenção de estar subsumido. (Agradecemos à Profa. Me. Cristina Gibk, pela valorosa discussão e interesse no tema nos gêneros textuais.)

⁴⁷ Abordagem de ensino de línguas postulada com base nos estudos funcionalistas de Michael Halliday (1973), que destaca o desenvolvimento da competência comunicativa como objetivo da aprendizagem (RICHARDS *et al.* 1992:65).

⁴⁸ Considerado um dos mais importantes antropólogos do séc. XX, Clifford Geertz foi o fundador de uma das vertentes da antropologia contemporânea - a chamada Antropologia Hermenêutica ou Simbólica ou Interpretativa -, que floresceu a partir dos anos 50. Sua tese principia na defesa do estudo de *quem as pessoas de determinada formação cultural acham que são, o que elas fazem e por que razões elas creem que fazem o que fazem*. Interessantemente para os propósitos da presente investigação, uma das metáforas preferidas, para Geertz, para definir o que faz a Antropologia Interpretativa é a da leitura das sociedades como textos ou como análogos a textos. (DURANTI, 1997)

textos que realizem com eficácia seus propósitos comunicativos, de acordo com o gênero a que pertencem” (BIASI-RODRIGUES *et al.*, 2009, p. 17).

Muitas investigações desenvolvidas no Brasil sobre ensino e aprendizagem de língua inglesa com finalidade acadêmica incorporaram o entendimento de Swales, que adota um ponto de vista sócio-retórico em relação ao tema, ao preconizar que “os gêneros são veículos de comunicação para atingir um objetivo” (1990, p. 46). Ou seja, mesmo sendo eventos linguísticos, os gêneros são concebidos aqui como atividades sócio-discursivas, uma vez que estão voltadas para a realização de ações comunicativas.

Para esse autor, o gênero de texto acadêmico “possui uma lógica própria porque serve a um propósito comunicativo em uma determinada comunidade discursiva (GIERING, ALVES e MELLO, 2010, p. 12)”. Há aqui duas premissas sobre as quais devemos refletir para melhor compreender as ideias de John Swales: os propósitos comunicativos e as peculiaridades da comunidade em foco.

É por meio de gêneros textuais que “as ações de linguagem dos indivíduos que se comunicam no contexto acadêmico se concretizam” (GIERING, ALVES e MELLO 2010, p. 7), variando de acordo com seus propósitos. O *abstract*, por exemplo, tem por finalidade sintetizar o texto de um trabalho acadêmico. Swales estabelece os propósitos comunicativos como sendo tanto um critério próprio ao gênero, quanto um critério que mantém a intenção desse gênero e que é compreendido através de uma ação retórica. Pode-se argumentar inclusive que os propósitos, moldados por critérios, são anteriores à estruturação esquemática do gênero mesmo.

Ao serem reconhecidos pelos membros aprendizes ou especialistas das comunidades que integram, os critérios, por sua vez, fornecem a razão de ser do gênero, contribuindo para a sua identificação, estruturação e convenções. Novos propósitos vão, portanto, delineando novos gêneros, ampliando a relação entre texto e contexto⁴⁹. Swales (1990), no entanto, ressalva que a identificação do propósito comunicativo nem sempre é fácil, porque há gêneros que são compostos por um conjunto de propósitos diferentes, a exemplo do *anúncio de uma prática no campo da saúde pública*, que pode, em determinados casos, informar, educar e entreter. Assim a ação sociorretórica nem sempre será determinante para a identificação de um dado gênero. Por isso, posteriormente, Swales (2004) acrescenta

⁴⁹ Esse entendimento se aproxima do aporte fornecido pelos estudos de gênero de uma corrente denominada *New Rhetorics* (MILLER, 1984), cuja ênfase está em estabelecer o papel social dos gêneros em detrimento de suas formas, linguagem típica ou organização interna, bem como valorizar o contexto a partir do qual se originam. Destaque-se que essa autora teceu uma das mais amplas análises de gênero como fenômeno social já conhecidas. Seus estudos partem da perspectiva retórica para definir gêneros como “ações retóricas típicas, baseadas em situações recorrentes” (p. 159). (“... *typified rhetorical actions based in recurrent situations*”).

dois procedimentos para o reconhecimento dos gêneros. São eles os procedimentos textuais e contextuais.

O procedimento textual trabalha com um conjunto de aspectos para definir o gênero, sendo eles os propósitos comunicativos, o levantamento da estrutura, estilo, o conteúdo e também o contexto de uso. Já no procedimento contextual, a análise se volta para a comunidade discursiva e seus valores, expectativas, condições, propósitos, características e normas de etiquetas, como é possível notar em relação à produção dos trabalhos acadêmicos. Swales (2004) estabelece que esses dois procedimentos se associam, para que seja possível aos membros de uma comunidade discursiva identificarem uma classe de eventos comunicativos.

As comunidades discursivas, por sua vez, são caracterizadas pelas práticas de linguagem adotadas nos eventos comunicativos que as definem. Quando os integrantes de uma comunidade dominam as convenções de um gênero que lhe é típico, são capazes de atingir os propósitos comunicativos implicados em suas ações retóricas com mais propriedade. Observe-se que aqui temos outro ponto de contato com Bakhtin. Enquanto aquele autor se referia a *esferas da atividade humana*, que, para ele, seriam todos os eventos em que se pode tomar parte, Swales toma por base a noção de comunidade discursiva que, mesmo podendo ser mais restrita, guarda semelhanças com a visão bakhtiniana em função de ambas se referirem a eventos de linguagem.

De acordo com Swales (2000), a comunidade discursiva apresenta as seguintes características que lhe são peculiares:

1. Conjuntos definidos de objetivos que podem ser estabelecidos pública e explicitamente e também podem ser, parcial ou integralmente, aceitos pelos membros;
2. Mecanismos de intercomunicação variáveis entre os membros para uma série de propósitos;
3. Capacidade para desenvolver gêneros característicos;
4. Aquisição e busca constante de léxico específico;
5. Membros com um grau adequado de conhecimento relevante e perícia discursiva.

Através do trabalho com os gêneros que lhe são característicos, a comunidade acadêmica dá conta de seus propósitos comunicativos. Para ser aceito, é preciso ter conhecimento das convenções que caracterizam esses gêneros. As experiências com essas práticas vão criando uma espécie de repertório de escolhas estratégicas dos modos de dizer dos produtores de textos desses gêneros acadêmicos, para que concretizem as metas de suas comunidades. O reconhecimento de tais metas fornece as bases, enquanto o desconhecimento das mesmas restringe a participação dos não-membros. “São essas escolhas textual-discursivas que constituem o *contorno retórico* dos textos” (GIERING, ALVES e MELLO, 2010, p.20).

O evento comunicativo é considerado como um conjunto formado pelo próprio discurso e pelos participantes, mas não somente, “também pelo papel do discurso e o ambiente de sua produção e recepção, incluindo suas associações culturais e históricas” (SWALES, 1990, p. 46).

Partindo do princípio segundo o qual essas associações culturais e históricas podem ser múltiplas, Swales também reconhece que um indivíduo não pertence a uma única comunidade discursiva, o que “acarreta variadas práticas sociais e diversos relacionamentos com outras comunidades” (BIASI-RODRIGUES, 2009, p. 24). Se as práticas então podem variar, é natural que surjam novos gêneros como fruto dessas práticas, bem como de seus propósitos, como discutido anteriormente.

Os gêneros se mantêm como uma rede interativa de comunicação através da qual os membros são avaliados de fora de seu contexto. Cumpre aos membros experientes o autoconhecimento sobre seus valores e identidades. Aos noviços, cumpre serem instruídos nas tradições e práticas da comunidade, para que se tornem membros plenos (SWALES, 1998). Por isso, para ser plenamente aceito em uma comunidade acadêmica, é preciso dominar além da *lingua franca* das publicações científicas contemporâneas, as formas de dizer a ciência, afinal nenhum conhecimento existe de forma autônoma, mas sim organizado pela linguagem. Nesse sentido, este estudo se volta para um contexto que envolve estudantes que precisam desenvolver articulação tanto para agir no subgrupo com que têm contato direto na universidade onde desenvolvem formação acadêmica, quanto no grupo maior de uma comunidade internacional de pesquisa. No entanto, nesse meio, encontram-se também os mestres que já atuam com destreza nessa comunidade maior e quem ter por objetivos alavancar a competência de novos usuários.

Enquanto os elementos linguísticos não são suficientes para a análise de um gênero, o contexto é elemento fundamental para a compreensão⁵⁰. Ao mesmo tempo, escritor e leitor precisam estar em sintonia quanto às convenções que dão legitimidade a cada gênero em uso na comunidade discursiva, constituída em torno de objetivos públicos comuns e de mecanismos de intercomunicação para o seu alcance. Quando os padrões de similaridade e os propósitos comunicativos se materializam em exemplares de um dado gênero, a comunidade discursiva os acolhe como prototípicos. Essa prototipicidade, no entanto, poderá ser relativa, como também reconhece o autor:

(...) o protótipo comunicativo foi nomeado como a propriedade privilegiada de um gênero. Outras propriedades, como forma, estrutura e audiência, operam para identificar até que ponto um exemplar é *prototípico* de um determinado gênero (SWALES, 1990, p. 52).

Por fim, uma outra característica endereçada por Swales é a terminologia adotada por uma comunidade discursiva, tanto para o uso quanto para nomear os gêneros que produz. “Os termos usados mostram como os membros percebem/entendem a ação retórica dos eventos comunicativos” (BIASI-RODRIGUES, 2009, p. 22). O autor, contudo, reconhece os aspectos problemáticos desse critério (SWALES, 1990), que ficam por conta da falta de univocidade; ou seja, o mesmo evento pode ser identificado por mais de um termo ou o termo pode não ser alterado, mas sim a atividade que o realiza. Esse aspecto é também bastante caro a este trabalho, uma vez que a investigação tem forte identificação com o campo da Terminologia de Perspectiva Textual. Nesse ponto, portanto, a análise entra em intersecção com aspectos que dizem respeito aos estudos dos textos especializados pela perspectiva dos estudos em Terminologia, já que estamos nos reportando ao gênero de texto artigo acadêmico, o qual se constitui a partir da atividade técnica ou científica.

1.2.3. Das contribuições de Douglas Biber a um ponto de confluência

A partir da década de 80, as contribuições de Douglas Biber para a Linguística Aplicada se voltam para a pesquisa de aspectos relacionados à caracterização das variedades textuais, cujo ensino o autor entende como a meta mais importante da educação formal. Nesse

⁵⁰ Novamente, percebe-se aqui um entendimento capaz de remontar aos postulados de Bakhtin.

contexto, afinal, os indivíduos são expostos a diferentes estruturas e padrões linguísticos que precisam reconhecer, interpretar e produzir.

Professor da Universidade do Norte do Arizona, Biber defende que o sucesso no ambiente acadêmico está atrelado ao aprendizado de padrões de linguagem determinados em função de situações e propósitos comunicativos específicos, embora essas situações e padrões possam variar também nesse contexto. Ele sustenta, portanto, o caráter variacionista⁵¹ da língua, que identifica também em relação às situações formais de uso⁵².

Esse linguista aplicado também reconhece que, antes de ingressarem na escola e aprenderem os gêneros acadêmicos, grande parte das crianças já domina outros tipos de narrativa, como a conversa e as histórias (BIBER e CONRAD, 2009, p. 3). Podemos então cotejar essa asserção ao ponto de vista de Bakhtin em relação à aprendizagem dos gêneros secundários. Além disso, é possível notar a presença de um princípio norteador em comum tanto com Swales quanto com Bakhtin, no que tange à variação do caráter prototípico de uma comunidade para outra. Biber sustenta que os fatores que impulsionam a variação nas escolhas linguísticas, além de sistemáticos, envolvem os propósitos de comunicação do falante, bem como as circunstâncias de produção (BIBER, 1995, p.1).

Outrossim, o estudioso destaca que a aprendizagem das diferenças entre os padrões de linguagem são ainda mais desafiantes para acadêmicos que precisam dominar uma outra língua. Afinal para manejar adequadamente uma variedade de texto⁵³, é preciso empregá-la com naturalidade. Ou seja, não é suficiente apenas memorizar o léxico e as convenções retóricas de uma língua; existem outros aspectos mais sutis que caracterizam o uso adequado dos textos, argumentação que também sustenta esta tese.

Biber vem dedicando uma parte significativa de seu trabalho para o estudo da sistematicidade dos padrões de uso da língua inglesa e para apontar como a Linguística de *Corpus* pode ser traduzida em aplicações para a sala de aula, de modo que se possa ensinar e aprender tais padrões. Uma de suas contribuições principais para a identificação dos padrões de uso da língua vem a ser em avançar nos estudos textuais a partir de seu entendimento de *registro*, pois, inicialmente, ele empregava apenas a noção de *gênero* (BIBER, 1988). Por *gênero*, correntemente, postula a descrição dos propósitos de uma variedade de texto com

⁵¹ Modo pelo qual as línguas variam, sistemática e coerentemente, de acordo com o contexto histórico, geográfico e sociocultural por influências dos usuários. Para conhecer o tema, vide Labov (2007).

⁵² Observe-se que a noção de língua com que estamos tratando aqui é estabelecida como dependente do contexto, sendo esse um aspecto central para o entendimento dos pressupostos biberianos.

⁵³ Quando Biber se refere a *texto*, ele quer abranger tanto manifestações orais quanto escritas de linguagem.

foco nas estruturas situacionais adotadas, a exemplo das convenções acerca de como uma carta começa e termina (BIBER e CONRAD, 2009, p.2).

Para Biber e Conrad (2009), os gêneros costumam ser governados por convenções específicas geralmente reconhecidas pelos membros de determinada cultura e nomeados por essa cultura (p. 34), o que se aproxima dos entendimentos de Swales e Bakhtin para gênero. A noção de *registro*, por sua vez, se organiza por associação entre propósito e contexto, para identificar as escolhas linguísticas de usuários proficientes em uma variedade textual. Ou seja, a perspectiva é estabelecida com base em restrições do léxico, enquanto o conceito de gênero se volta para o objetivo do discurso. No entanto, ambas as noções se baseiam no contexto situacional e não subsistem de modo independente portanto. Por esse motivo, Biber costuma se referir muitas vezes a ‘gênero/registo’ de forma conjunta, pois ele também define registro como “variedades definidas situacionalmente, descritas a partir de características léxico-gramaticais peculiares” (2006, p.11)⁵⁴, que incluem, além do propósito da comunicação e das circunstâncias de produção, a forma de ocorrência – escrita ou oral - e características demográficas do falante ou autor (BIBER, 1999, p.5)⁵⁵.

Essa combinação entre a análise de características linguísticas comuns a uma variedade de textos e a análise de situação de uso se torna central em suas pesquisas, embora a distinção entre os dois lados da moeda que distinguem gênero e registro não tenha ficado clara desde o princípio das teorizações de Biber. “Os registros compartilham várias características linguísticas - como substantivos, pronomes, verbos, adjetivos, etc. – e eles se distinguem pelo uso relacionado a essas características”⁵⁶ (BIBER *et al.*, 2000, p. 136). Desse modo, textos situacional ou linguisticamente semelhantes representam um registro. A asserção subjacente é de que itens linguísticos como pronomes e verbos são adotados funcionalmente em associação aos propósitos comunicativos e contextos situacionais dos textos. As escolhas dos itens léxico-gramaticais tendem a ocorrer porque esses *funcionam* em relação aos propósitos e ao contexto situacional de um registro específico.

Biber (1999) identificou quatro tipos básicos de registros: a conversa cotidiana, a ficção, a linguagem jornalística e a prosa acadêmica, que estabelece como podendo ser subdivididos, como, por exemplo, um jornal compreende textos de reportagem e editoriais.

⁵⁴ “*Situationally defined characteristic lexico-grammatical features.*” (BIBER, 2006)

⁵⁵ Embora Biber faça amplo uso da noção de registro em suas investigações, observemos que Swales (1993^a) já mencionava que “os registros impõem limitações nos níveis linguísticos de vocabulário e sintaxe, enquanto as limitações dos gêneros operam no nível da estrutura do discurso” (p.41). “Registers impose constraints at the linguistic levels of vocabulary and syntax, whereas genre constraints operate at the level of discourse structure.” (SWALES, 1993, p.1)

⁵⁶ “*Registers share many linguistic features – such as nouns, pronouns, verbs, adjectives, etc. – and they are distinguished by the relative use of these features*” (BIBER *et al.*, 2000, p. 136).

Um registro, da mesma forma como um gênero, se define em relação ao público a que se destina. O registro acadêmico, por exemplo, revela um menor uso de pronomes pessoais em primeira e segunda pessoas do que o registro jornalístico, assim como revela o uso de itens lexicais prototípicos que este estudo busca identificar e organizar. Biber assevera que a perspectiva de registro se presta para descrever todos os tipos de variedades textuais, no entanto, nem todos são reconhecidos do mesmo modo como o são os gêneros.

Por exemplo, “uma conversa casual entre colegas” descreve uma variedade de texto que ocorre em uma situação particular de uso, e há características linguísticas funcionalmente associadas com essa situação. Portanto, essa variedade pode ser analisada pela perspectiva de registro, embora não represente um gênero claramente definido. (BIBER e CONRAD, 2009, p. 34)⁵⁷

Ao mesmo tempo, em termos de graus de especialização de uso de variedades textuais, Biber e Conrad (*op. cit.*) também ponderam que, em alguns casos, domínios especializados possuem múltiplos gêneros/registros que vêm a ser familiares aos especialistas em um campo, mas não ao público em geral (p. 34), como no caso da distinção entre os *abstracts* indicativos, informativos e críticos, em contexto de escrita acadêmica⁵⁸. Frequentemente, os membros especializados de uma determinada cultura estabelecem essa diferenciação a partir dos propósitos comunicativos desses textos, embora as circunstâncias possam parecer quase idênticas aos olhos de um observador casual. Desse modo, além de perceber diversidades situacionais para os usos de variedades textuais, Biber também enfatiza o caráter cultural que estabelece essas diferenças. Ou seja, os textos acadêmicos não detêm características universais, como veremos mais adiante, tampouco apresentam registros idênticos de uma área de conhecimento para outra em uma mesma universidade.

Um ponto a destacar acerca da trajetória investigativa de Douglas Biber é que, apesar de as ideias desse pesquisador serem correntemente adotadas por boa parte dos estudos em variação textual em Linguística de *Corpus*, ele não iniciou suas investidas acadêmicas originalmente como um linguista de *corpus*. Foi na década de 90 que Biber passou a se definir como tal e desde então se destaca como representante do campo nos Estados Unidos. Essa questão é relevante, porque os estudos linguísticos com base em gênero costumam descrever a organização dos textos com relação a aspectos retóricos, sem preocupação específica com a

⁵⁷ “For example, “casual conversation among colleagues” describes a text variety that occurs in a particular situation of use, and there are linguistic features functionally associated with that situation. Thus, this variety can be analyzed from a register perspective, even though it does not represent a well-defined genre.” (BIBER e CONRAD, 2009, p. 34)

⁵⁸ Distinções que serão abordadas no cap. 3.

frequência de ocorrência de itens lexicais. Embora, a propósito, Swales (1990; 1992; 1998) também tenha enfatizado o conteúdo - assim como o estilo - a partir da razão de ser de um gênero discursivo definido por uma dada comunidade, o item frequência de ocorrência de itens léxico-gramaticais não é uma de suas preocupações maiores.

Já para Biber, a introdução da noção de registro abre o escopo das investigações textuais para agregar o indiciador frequência como determinante da caracterização da prosa acadêmica. Isso porque ele enfatiza o caráter recorrente de registro, que pode ser identificado como comum a uma variedade acadêmica, por exemplo, porque ocorre em excertos ou em textos completos. Ao mesmo tempo, Biber destaca que a mesma característica não se verifica necessariamente em relação aos gêneros, tendo em vista que as variedades situacionais que os determinam podem acontecer uma única vez ao longo de um texto, como, por exemplo, uma resenha acadêmica pode ser identificada pelas formalidades de abertura do texto.

Por fim, em suas pesquisas, Biber examina as estruturas léxico-gramaticais de uma quantidade de textos de diferentes registros, especialmente o acadêmico, e se concentra no uso das particularidades encontradas nas diferentes variedades da língua inglesa (BIBER *et al.*, 1999, p. 4), a fim de descrever a organização e as distinções entre os registros a partir desses aspectos de uso em comum. Pode-se dizer que, para dar conta dessa empreitada, esse estudioso, na verdade, aprofundou o entendimento contemporâneo de variedade textual a partir da conjugação entre *registro* e *gênero* com base na perspectiva do uso situacional da língua.

Antes de passarmos ao próximo capítulo, cabe reforçar que as ideias de Douglas Biber são concebidas aqui como um ponto de confluência dentre os três pesquisadores apresentados. Isso porque defendo que se pode ler, nas concepções desse último autor, uma conjugação das perspectivas de Bakhtin e Swales. Estabelecem-se esses pontos de união a partir de vários aspectos. Biber assume a ótica variacionista, o que implica reconhecer a influência sociocultural dos usuários e aceitar o caráter plástico do discurso. Ele incorpora indiretamente a posição enunciativa para os textos acadêmicos, ao se apropriar das noções de *gênero* e *registro*. Ele também parece conciliar os entendimentos dos dois outros linguistas, quando assevera que os indivíduos são expostos a diferentes estruturas e padrões linguísticos, que precisam reconhecer, interpretar e produzir. Biber, igualmente, estabelece a ideia de sistematicidade que integra os *tipos relativamente estáveis*, de Bakhtin, à *estrutura esquemática do discurso reconhecida pelos membros de uma comunidade*, de Swales. Para

encerrar, ele parece agregar seu próprio tempero ao incluir uma preocupação com a frequência de ocorrência dos itens lexicais.

Se concordamos que é o ponto de vista que faz o objeto, como nos ensinou Saussure (2006), é a perspectiva do uso social e, portanto, flexível, da linguagem que está em pauta na conjugação dessas perspectivas. Compreendo que a soma dessas ideias confere vigor e substância para o estudo do objeto em questão, o *abstract*, classe de eventos comunicativos relativamente estáveis e situacionais – para conjugar os entendimentos dos três autores da análise.

No próximo capítulo, abordaremos a constituição do texto especializado pelo olhar dos estudos em Terminologia. Para tanto, busca-se mapear as características através das quais o texto especializado se apresenta para o nosso reconhecimento e como essa forma de registro se torna instrumento de participação no mundo do conhecimento acadêmico.

2. CAPÍTULO 2

“...o texto especializado também se mostra como fruto da ação perceptiva e transformadora de um sujeito enunciador, individual e múltiplo, sobre um conjunto de conhecimentos e textos com os quais se relaciona. Essa ação do sujeito que produz o texto pode ser vista como redizer algo ou recontar a estruturação de um conhecimento, tornando-o acessível ao leitor.”

KRIEGER e FINATTO, 2004, p. 194

O TEXTO ESPECIALIZADO COMO FORMA DE PARTICIPAÇÃO EM UM MUNDO QUE SE COMPLEXIFICA

As grandes transformações ocorridas a partir do século XX, por conta da modernização tecnológica e do avanço no campo das comunicações, desencadearam não só uma evolução nos modos de viver, mas igualmente nos modos de nomear o mundo. Consequência natural dessa evolução, a Terminologia, área de saber multidisciplinar, amparada principalmente na dinamicidade denominativa e conceitual das línguas, volta-se para o estudo dos fenômenos da comunicação especializada.

Com o intuito de chegar, especificamente, até ao estudo das linguagens especializadas, que se manifestam por meio dos textos técnico-científicos, trataremos primeiro de contextualizá-las a partir do campo mais amplo que lhes dá origem. Linguagens especializadas, em que pese saber, “são linguagens de grupos ou linguagens especiais (socioletos), as quais estão caracterizadas por usos linguísticos de determinados grupos profissionais e, portanto, também constituídas por estratos sociais” (HOFFMANN, 2015, p. 42).

2.1. Jovem ciência do léxico

A Terminologia, no âmbito do Brasil, é a mais jovem das ciências do léxico e se estabelece tendo por objeto principal os termos técnico-científicos. Como veremos a seguir, no entanto, as novas vertentes desse campo tenderam a se ocupar de elementos mais amplos do que a unidade lexical terminológica.

Não obstante, o uso de termos relacionados ao conhecimento especializado possui registros que, do mesmo modo como o estudo dos textos, remontam à Antiguidade Clássica. Esses se baseavam na compilação e investigação de “palavras especializadas” dos diferentes campos de conhecimento.

Nos séculos XVII e XVIII, a área passa a ser concebida de modo coerente a partir da unificação e regulamentação do léxico de diferentes disciplinas científicas, como a Botânica e a Química. É com base na necessidade de precisão denominativa e conceitual que os cientistas desse período padronizaram terminologias, a fim de facilitar a comunicação profissional e a transferência de conhecimentos.

Essa demanda por unificação faz surgir, formal e institucionalmente, no século XX, um “campo de estudos teóricos e aplicados que se ocupa dos termos científicos e técnicos com distintos focos de interesse e de perspectivas” (KRIEGER, 2011, p. 445). De modo semelhante à evolução dos estudos textuais, a Terminologia dos séculos XX e XXI também se expandiria por caminhos e perspectivas diversificados.

Para Rondeau (1994), a constituição da área propriamente como disciplina científica se deu a partir da Escola Russa, em 1930, com o estabelecimento de uma reflexão teórica sobre princípios, métodos, funcionamento e características das línguas de especialidade e dos termos, postulada por D.S. Lotte. As investigações desse engenheiro, que analisava os fenômenos terminológicos, como sinonímia e polissemia, eram de caráter descritivo, mas também normativo.

Para muitos, no entanto, a exemplo de Felber e Picht (1984), a Terminologia só ganhou estatuto de área de investigações com os estudos feitos na Áustria, também a partir da década de 1930, quando nasce a Teoria Geral da Terminologia (TGT). A TGT se estabeleceu a partir das preocupações seminais do engenheiro Eugen Wüster com a normatização dos termos técnicos. Wüster, fundador da chamada *Escola de Viena* de estudos do campo, imbuíu-se do espírito de inovação tecnológica em que se encontrava a Alemanha nos anos 20, para dar luz ao princípio da univocidade entre o conceito e o termo que o designa. A concepção era de que um único termo deveria designar um conceito. O conceito, por sua vez, como conhecimento, precederia a expressão. Esse investigador acreditava que assim seriam dirimidos os ruídos de comunicação entre os especialistas e que a ambiguidade seria eliminada nos discursos técnicos e científicos⁵⁹. A TGT, estabelecida postumamente em livro organizado por alunos de Wüster, dissociava léxico e gramática, apresentando-os como unidades independentes.

Na análise de Zilio (2011), o entendimento da TGT desprezaria o contexto por não levar em consideração questões linguísticas de extrema importância para distinguir *termos* de *palavras*. Cotejados aos postulados de Sausurre (2006), por exemplo, Zilio (OP. CIT.) reflete que

⁵⁹ Observe-se ser este um dos pontos de crítica de Swales, descrito no capítulo anterior (p. 20), justamente em relação à terminologia adotada pelos membros de uma comunidade discursiva. Swales se refere especificamente às ações retóricas dos eventos comunicativos especializados, uma vez que, para ele, o mesmo evento pode ser identificado por mais de um termo ou o termo pode não ser alterado, mas sim a atividade que o realiza.

Se entendemos que a significação de uma unidade se dá no eixo sintagmático, também entendemos que um termo somente se apresenta como tal nesse eixo. Um termo não existe, portanto, fora de seu contexto linguístico. Em outras palavras, um termo não o é. São as palavras ou sintagmas da língua que adquirem no eixo sintagmático o valor, ou a significação, de termo. (ZILIO, 2011, p. 125)

Ademais, concordo com Zilio (*op. cit.*), quando esse pesquisador assevera que, distinguir termos de palavras, de todo modo, não parece tarefa útil, uma vez que “todas as palavras seriam termos ou todas seriam palavras da língua comum” (p.125).

Ainda a respeito dessa diferenciação, Cabré (2013), do mesmo modo, não estabelece mais separação apriorística entre palavra e termo. A estudiosa defende que o movimento de *ativação terminológica*, segundo o qual o termo é um signo poliédrico e não apenas um elemento cognitivo como concebido pela teoria clássica da TGT, encontra contemporaneamente espaço e condições explicativas no âmbito da epistemologia da ciência da linguagem.

Não obstante, Wüster foi, na verdade, o desbravador do campo no ocidente. Nos estudos desse pesquisador, a Terminologia é concebida como um segmento interdisciplinar do qual a Linguística é um dos polos de convergência, ao lado das ciências cognitivas, da comunicação e da informática, privilegiando aspectos cognitivos e normativos das terminologias. Krieger (2011) descreve essa primeira fase da Terminologia como “vocacionada para a problemática de padronização internacional dos termos técnico-científicos, privilegiando assim a missão de controlar os usos terminológicos no plano mundial” (p. 446).

Acerca dos períodos concomitantes de criação das Escolas Russa e Austríaca, poderíamos intuir mesmo que as demandas estabelecidas a partir da evolução do conhecimento especializado, nas primeiras décadas do século passado, encarregaram-se dessa coincidência. Afinal os dois estudiosos não desenvolveram trabalhos conjuntos ou sob influência mútua.

As décadas que se seguiram à apresentação das pesquisas de Wüster, contaram com investigações de caráter linguístico. Essas investigações logo identificariam a insuficiência do modelo wüsteriano para lidar com a análise da unidade terminológica enquanto signo linguístico composto, ao mesmo tempo, por conteúdo e expressão. Constatou-se igualmente a inadequação da TGT para a análise terminológica cotejada a outros aspectos do texto e da comunicação especializada. Uma nova linha de raciocínio surgiria, e o conceito passaria a ser concebido a partir do significado, abrangendo, por extensão, elementos

pragmáticos. Nas palavras de Barros (2006), “o signo terminológico passou a ser tratado de acordo com a concepção saussureana de unidade entre o significante e o significado”.

Em um rastreamento cronológico da evolução do campo a partir dessa virada, os anos 80, veriam surgir, no Canadá, a Socioterminologia, amparada nos campos da Sociolinguística e da Etnografia. Para essa vertente, são os diferentes usos que a comunidade faz dos termos que levam ao surgimento e emprego de variantes terminológicas, a exemplo da linguagem da aviação. O postulado máximo da Socioterminologia é “ter na base da pesquisa a variação linguística dos termos no meio social e, por consequência, entender a mudança terminológica como mecanismo resultante da pragmática discursiva” (FAULSTICH, 2006, p. 29). Uma das grandes contribuições sob influência dessa linha de pesquisa é estabelecimento da ISO (*International Standardization Organization*), pela Comissão de Nomenclatura Anatômica Internacional.⁶⁰

O início da década de 1990 testemunha o florescimento da Teoria Comunicativa da Terminologia (TCT), calcada nos estudos coordenados pelo grupo de Maria Teresa Cabré, na Universidade Pompeu Fabra, em Barcelona, na Espanha. A TCT se revela uma abordagem de “base linguística que zela pelo caráter comunicativo dos termos inseridos na linguagem efetivamente utilizada em ambientes específicos” (SALES, 2007, p.2). O objeto de estudo da TCT são as unidades terminológicas utilizadas em áreas de conhecimento especializado e adotadas com vistas à comunicação. Para essa vertente, qualquer processo de comunicação envolve variação de formas lexicais, que se manifestam como denominações alternativas para um mesmo conceito (sinonímia) ou pela abertura semântica de uma forma (polissemia). Os termos passam a ser contemplados como de natureza poliédrica, pois integram aspectos linguísticos, cognitivos e sociais, que podem ser estudados de modo conjunto ou independente. Cabré estabelece a *Teoria das Portas*, que abre a possibilidade dos termos serem estudados por perspectivas variadas, a partir das ciências linguísticas, cognitivas e da comunicação.

As contribuições da TCT para o campo são representadas pela prevalência das expressões à frente dos conceitos. Além disso, essa corrente percebe o conhecimento técnico-científico como situado, em oposição a ser detentor de uma condição universal. Do mesmo modo, a TCT também se caracteriza pela noção de que os âmbitos especializados não são neutros e tampouco consensuais.⁶¹

⁶⁰ Vide Faulstich (2006), para conhecer os trabalhos do campo e as pesquisas no Brasil.

⁶¹ Vide Cabré (1999; 2002) para aprofundamento das pesquisas em TCT.

O final da década de 1990, trouxe a Terminologia Sociocognitivista (TS), originária da Europa e centrada nas pesquisas de enfoque cognitivista de Rita Temmerman, na Bélgica. Essa linha, baseada em princípios da Semântica Cognitiva e da Sociolinguística, procura integrar aspectos sociais, culturais e a cognição, para explicar como acontecem os processos cognitivos e buscar mostrar as relações entre as dinâmicas de categorização e a linguagem. A proposição é de que se fale em *unidades de compreensão*, em vez de *conceitos*, pois tais unidades estão em constante evolução. O texto passa a adquirir as dimensões pragmática, linguística e comunicativa, além da linguística. Outro diferencial de abordagem da TS é a valorização de elementos diacrônicos para a descrição e explicação de toda uma série de fenômenos terminológicos.

A TS estabelece a importância do campo ser estudado de maneira mais sistemática, através do desenvolvimento de métodos para a aproximação do conjunto de aspectos que fazem parte do entendimento das linguagens especializadas. As investigações de se voltam para explorações terminológicas em língua inglesa, por ser essa a língua da ciência e da tecnologia, ou a língua de formação mais primária, no dizer de Sager (1993)⁶².

O começo do século XXI veria surgir outras novas tendências dos estudos de Terminologia. Entre elas, a Terminologia Cultural (TC), em torno das ideias de Marcel Diki-Kidiri (2002), no continente africano. Essa vertente propõe que se leve em conta, para o estudos das terminologias e sua implantação ou gerenciamento, todo um contexto da diversidade cultural e linguística nas sociedades, tomando como exemplo o que ocorre nas sociedades africanas. Essa perspectiva apresenta como base a premissa de que a apreensão da realidade e também as denominações utilizadas para expressar o conhecimento especializado estão vinculadas à percepção cultural dos integrantes de uma comunidade, à semelhança do que Temmerman (2000) já havia proposto. Rompendo com tendências formalistas, o entendimento básico aqui é de que os conceitos não são universais. Nessa abordagem, os falantes de uma dada língua, em meio a uma dada cultura, são levados a criar e a adotar termos que tenham a ver com a sua realidade. Para Diki-Kidiri (*op. cit.*), o funcionamento das terminologias que se fixam em determinados contextos comunicativos, como os de países periféricos e/ou em desenvolvimento, vai além do uso de termos meramente impostos pela cultura dominante.

Concepções inovadoras como as de Diki-Kidiri corroboram o fato de que, mesmo sendo um campo de estudos recente, a Terminologia se mostra epistemologicamente

⁶² Vide Temmerman (1997; 2000) para conhecer os trabalhos do campo.

robusta o suficiente para repensar radicalmente as bases que a constituíram. O intuito é explorar novas possibilidades conceituais e aplicadas em um mundo cujas denominações especializadas estejam submetidas aos seres humanos e não vice-versa⁶³.

Por fim, a complexidade dos textos especializados também motivou a criação de uma corrente de estudos especialmente atrelada à noção de texto para os estudos de Terminologia. Estabelece-se, então, um campo que vem sendo fruto de várias denominações. Alguns se reportam a ele como Terminologia Textual, outros como Linguística do Texto Especializado, Linguística das Linguagens Especializadas (LLE)⁶⁴ (BOURIGAULT e SLODZIAN, 1999; HOFFMANN, 1988^a; 1988^b; FINATTO, 2004; CONDAMINES, 2005; KRIEGER, 2008; ZILIO, 2012^b) ou ainda como estudos em Terminologia de Perspectiva Textual (CIAPUSCIO, 2003). Finatto (2007) assevera que “nessa perspectiva, a porta de entrada para o edifício do texto é a sua superfície e a sua estrutura concretamente expressa, embora não se negue que haja outras possibilidades de acesso, outras portas” (p. 450).

Essa tendência das investigações de Terminologia, com a qual se identifica este trabalho e que ainda está começando no Brasil, tem sido, outrossim, fonte de algumas controvérsias e estranhamentos. Esses advêm da ideia de uma Linguística das Linguagens Especializadas⁶⁵ ou Linguística do Texto Especializado, que seriam, em tese, áreas distintas da Terminologia.

A problematização se deve especialmente ao fato de que, como nos aponta Zilio (2010), a palavra *Terminologia* remete ao estudo do *termo* e, portanto, parece estranho que algo terminológico possa também se voltar, preponderantemente, para o aspecto textual. Ademais, alguns pesquisadores, a exemplo de Condamines e Krieger, assumem que o estudo e a identificação dos termos seja ainda o objeto de estudo central do campo⁶⁶. Defende-se aqui que o projeto epistemológico da área é bem mais amplo do que esse, como já estabeleceu Ciapuscio:

[...] a ideia básica é a de que o estudo terminológico do texto pode trazer contribuições não apenas para a terminologia, mas também para o esclarecimento de problemáticas relevantes para os estudos textuais, como, por exemplo, avanços

⁶³ Vide Diki-Kidiri (2002) para detalhamento das bases da Terminologia Cultural.

⁶⁴ Nos termos deste estudo, assume-se esta denominação, cujos estudos serão explorados mais detidamente na próxima seção.

⁶⁵ “A Linguística das Linguagens Especializadas proporciona uma das justificativas mais essenciais a favor da necessidade de que haja formação específica em linguagens especializadas e, ao mesmo tempo, fornece o material linguístico para esse tipo de formação.” (HOFFMANN, 2015, p. 42-3).

⁶⁶ Vide Zilio (2010) para aprofundamento dessa discussão.

em direção a uma tipologia dos textos especiais e ao estabelecimento de graus de especialização dos textos tomando como base o léxico (CIAPUSCIO, 2003, p. 85).

Desse modo, pela ótica dos estudos de Ciapuscio e Hoffmann, esse aporte teórico também representa uma evolução significativa nas pesquisas do campo ao se estabelecer não mais centrado no estudo dos *termos no texto*, mas agregando outros aspectos do texto especializado. Passa-se então a propor uma concepção dos *textos com termos*. A partir disso, um número de pesquisadores que se filiam a essa concepção (a exemplo de HOFFMANN, 1998; ZILIO, 2010, 2012^b e FINATTO, 2004, 2007, 2010, 2011^a, 2011^b) acreditam que compreender os modos de dizer da ciência e do conhecimento especializado naturalmente perpassa o reconhecimento das suas terminologias, mas pode ir além dessas, a fim de construir conjuntos de forma e significados ainda mais amplos.

Nessa direção, para as investigações dos textos que contêm termos, a linguagem é um sistema que se atualiza em textos, sendo esses, do ponto de vista comunicativo, “o signo linguístico primário” (HOFFMANN, 2015, p.47). Observe-se que o conceito de texto especializado, neste momento, compreende o texto além da dimensão escrita, assim como o léxico especializado vai além da mera adoção de nomes para designar os conceitos. Esses estudos fundamentaram a importância do diálogo entre a Terminologia e a Linguística Textual para explicar textos especializados.

A breve história do campo da Terminologia, aqui resumida, vem apontando, a partir do final do século XX, proposições que representam novas concepções epistemológicas em relação às orientações de seu fundador austríaco. Mais recentemente a área vem experimentando a passagem de um estado normativo para um estado descritivo, além do reconhecimento da necessidade de padronizações. Além disso, “assume uma face linguística como um campo de conhecimento, cujos parâmetros epistemológicos se situam efetivamente no âmbito das ciências da linguagem” (KRIEGER, 2011, p. 447). Esses novos entendimentos trabalham em defesa de aceitar-se que a distinção entre conhecimento *in vitro* (aceito por consenso ou padronizado) e conhecimento *in vivo* (espontâneo ou natural) determina toda a diferença metodológica, para dar conta de situações autênticas de uso da linguagem. Isso porque o primeiro é prescritivo, já que a variação é um aspecto controlado, e o segundo descritivo. Ainda, se o *termo* era concebido até a década de 80 como peça independente e dominante nas investigações de Terminologia, nos últimos anos, o *texto especializado* vem marcando seu espaço, como “o habitat natural das terminologias” (KRIEGER, 2011).

É relevante também ressaltar a importância do campo para o ensino de *English for Academic Purposes* (EAP) e mesmo para a organização de materiais e ferramentas didáticas. A vertente especializada da língua e, por extensão, o léxico específico de áreas de conhecimento aplicadas são a razão de ser da abordagem. Tornam-se cada vez mais frequentes as publicações de livros, dicionários e glossários específicos para determinadas áreas de conhecimento e é comum nos depararmos com títulos como Inglês para Medicina ou para Informática, Engenharia Elétrica (a exemplo de FADANELLI e FINATTO, 2015), Hotelaria, etc. Os estudos em Terminologia constituem-se em grandes aliados dos trabalhos em ensino e aprendizagem da língua inglesa, tendo em vista que é preciso desenvolver “competência linguística para agir na especialidade” (BUHLMANN/FEARNS, 1987, P.87).

2.2. Dando forma ao léxico da ciência

A partir dos anos oitenta e noventa do século XX, cresce o interesse dos pesquisadores da linguagem pelas formas de comunicação que caracterizam o conhecimento especializado. Para se chegar a este, é preciso desenvolver referências especializadas ou tecnológicas novas, as quais se organizam e se estabelecem através de uma linguagem específica.

Enquanto o conhecimento geral é conceitualizado através da experiência e do contato com o mundo real, o conhecimento especializado é descrito por intermédio de esquemas de referência pré-estabelecidos. Isso não significa dizer que esses mundos estejam em oposição, já que o conhecimento especializado também faz parte do *mundo real*. Mas esse tipo de conhecimento, de certo modo, está para o mundo como os gêneros secundários estão para os primários, conforme discutimos acerca dos postulados de Bakhtin (1997), já que precisa ser aprendido pelo contexto da técnica ou da ciência.

De forma análoga, embora o texto através do qual a linguagem especializada circula faça uso de recursos específicos que lhe dão um caráter determinador de sentidos discursivos, ele também pertence à linguagem natural adotada para a comunicação não especializada. Ou seja, as linguagens especializadas não estão em oposição à língua comum, mas fazem parte dela ao pertencerem às linguagens naturais, porém determinadas pela temática e pela área, bem como pela frequência de ocorrência.

Dito de outro modo e por uma outra perspectiva que se pode entender como

semelhante aos estudos de textos especializados, recorreremos aos postulados de Douglas Biber acerca de registro, para ilustrar como a língua natural está para a comunicação especializada. Para Biber e Conrad (2009), a análise linguística de um registro se baseia em palavras ou características lexicais que são tanto (1) recorrentes quanto (2) frequentes, porque ocorrem mais comumente nos textos de um dado registro do que de em outro. No entanto, aspectos que caracterizam um registro não se restringem a ele, a exemplo do uso de voz passiva (*was based on*), construção típica de textos acadêmicos que pode ser encontrada em todo registro, mas que, no entanto, é muito mais comum em escrita acadêmica do que em outros registros (p. 53). Assim, tomamos emprestados recursos da linguagem natural, para integrar a linguagem especializada e, com isso, dar forma ao léxico da ciência.

Beaugrande (1997), referido anteriormente, no panorama dos estudos em Linguística Textual, também destacaria que a comunicação da ciência, em sua gênese, *é feita da mesma matéria* que um texto não especializado. Ele observa que é preciso ter todo o domínio do discurso, a fim de fomentar o acesso ao conhecimento, não bastando apenas conhecer alguns de seus traços. Esse posicionamento reforça o enquadramento estabelecido para este estudo, de que a fluência está relacionada tanto a aspectos discursivos de caráter amplo, quanto às escolhas léxico-gramaticais que cabem ao aprendiz desenvolver com propriedade e ao professor e/ou ao pesquisador aplicado investigar e encontrar modos de organizar.

Cumprir notar, outrossim, que os pesquisadores voltados para o estudo do texto especializado têm adotado denominações heterogêneas para nomear a linguagem que lhe é característica, assim como apresentam entendimentos distintos para sua constituição. Enquanto Hoffmann (1998) emprega o termo *linguagem especializada* e adota o princípio de que o texto que lhe dá origem não pode ter sua base relacionada exclusivamente ao elemento temático das áreas de conhecimento, Schröder (1991), assim como Cabré (1999), lançam não da denominação *comunicação especializada*, que esta última autora relaciona especificamente à temática. Alguns linguistas aplicados, no entanto, referem-se a *línguas de especialidade*⁶⁷. Estes últimos investigadores são aqueles orientados para os domínios de *LSP (Language for Specific Purposes)*⁶⁸, área que Krieger e Finatto (2004) esclarecem que

⁶⁷ Este texto se afilia à denominação *linguagem especializada*.

⁶⁸ Ou, no caso específico da língua inglesa, *English for Specific Purposes (ESP)*. Krieger e Finatto (2004) asseveram que tais estudos, por sua vez, “não se confundem com os da Terminologia, posto que a origem de seu desenvolvimento está relacionada a um enfoque didático, voltado à instrumentalização da aprendizagem em línguas estrangeiras” (p. 115).

desenvolve-se fundamentada na concepção da existência de uma língua particular com autonomia própria, com particularidades e traços distintos do sistema linguístico por meio do qual a informação profissional é veiculada. Em razão dessa configuração, só poderia ser empregada por iniciados de uma matéria, únicos a entender os jargões nela utilizados (KRIEGER e FINATTO, 2004, p. 115).

Ressalve-se que essa noção de *língua de especialidade* a partir da qual se organiza a área de *LSP* e que é compreendida como “um subsistema linguístico particular e distinto do sistema geral das diferentes línguas naturais” (KRIEGER & FINATTO, 2004, p. 115) consiste em um entendimento controverso dentre os pesquisadores que adotam uma visão linguística das terminologias. Tal concepção se opõe ao pressuposto que apresentado acima, segundo o qual as linguagens especializadas não estão em oposição à língua comum, mas antes fazem parte desta.

Um dos principais pesquisadores interessados em identificar funções e critérios da linguagem especializada a partir de uma perspectiva linguística é Lothar Hoffmann, professor aposentado da Universidade de Leipzig. Seu ponto de vista é de que, ao adquirirem o *status* de linguagens de grupos, a função simbólica das linguagens especializadas se associa à função sintomática. Esse aspecto contribui “por um lado, para separar grupos de especialistas de outros grupos de pessoas, além de separá-los entre si, e, por outro, para unir os participantes do grupo” (HOFFMANN, 2015, p. 80). É desse modo que surge a identidade linguística dos grupos, que podem ser mais ou menos especializados e, com isso, mais ou menos empoderados pelos seus pares em relação ao domínio do conhecimento especializado, a exemplo dos autores experientes dos gêneros acadêmicos em relação aos novatos no campo.

Quanto à composição dos textos especializados, segundo Hoffmann,

O texto especializado é o instrumento ou o resultado de uma atividade comunicativa socioproductiva especializada. Compõe uma unidade estrutural e funcional (um todo) e está formado por um conjunto ordenado e finito de orações coerentes, pragmática, sintática e semanticamente ou de unidades com valor de oração que, como signos linguísticos complexos do conhecimento humano e das circunstâncias complexas, correspondem à realidade objetiva (HOFFMANN, 1998, p. 77).

A asserção declara o entendimento do autor de que o texto especializado não se resume aos termos científicos e técnicos, referido anteriormente, pois envolve todos os recursos linguísticos empregados em um âmbito de comunicação especializada para garantir a compreensão entre os usuários.

Finatto igualmente reforça que o texto especializado representa “um todo de sentido e de comunicação, constituído pelo conteúdo exposto e também pelo quadro geral das escolhas ou convenções do seu ‘modo de dizer’” (FINATTO, 2012, p. 156). Assim, o componente principal de suas preocupações é o “conjunto de elementos sintáticos, semânticos e pragmáticos que constituem as práticas textuais técnico-científicas” (FINATTO *et al.*, 2010, p. 154). Esse foco de atenção abrange o estudo do uso e dos padrões de uso, bem como das condições pragmáticas da comunicação, aspectos esses eleitos como base analítica para esta investigação.

Hoffmann salienta que “a especificidade das linguagens especializadas se expressa principalmente pela frequência de uso de determinados recursos linguísticos comprováveis com o auxílio do método da Linguística Estatística” (HOFFMANN, 2015, p. 41), sem perder de vista os questionamentos qualitativos. Esse ponto nos remete à Linguística de *Corpus* e nos aproxima novamente de Douglas Biber e sua preocupação com a frequência de ocorrência de itens lexicais. Até mesmo porque o foco de atenção em frequência para a LLE abrange o estudo do uso e dos padrões de uso a partir de, entre outros aspectos, combinatórias de palavras mais ou menos fixas, elementos retóricos e argumentativos, colocações especializadas, condições pragmáticas da comunicação e condições do gênero textual. Tais aspectos se relacionam à unidade de análise do estudo.

Notadamente, Hoffmann (2015) reforça, desde longa data, que os textos especializados apresentam, em média, sessenta por cento 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 (p. 43-4). “Também é comum que apenas substantivos sejam arrolados a uma terminologia, possivelmente adjetivados por outros substantivos ou adjetivos, ainda que se observe uma tendência à terminologização de verbos especializados.”⁶⁹ (HOFFMAN, 2015, p. 96). Isso porque “é preciso designar a multiplicidade de objetos e manifestações, que caracterizam a atividade especializada” (HOFFMANN, 2015, p. 43).

A esse respeito, Finatto e Kilian (FINATTO E ZILIO, 2015) mencionam que seria “interessante nos perguntarmos, hoje, se essa proporção se confirmaria em diferentes línguas e linguagens especializadas” (p. 38), o que vem a ser um dos objetivos deste estudo, voltado para língua inglesa.

As linguagens especializadas, por sua vez, apresentam restrições, que

⁶⁹ Os estudos de Biber, a propósito, também corroboram a asserção de que uma das principais características da prosa acadêmica contemporânea é o uso majoritário de nominalização (BIBER e GRAY, 2013⁹, p. 1).

Hoffmann (2015) aponta com base nos textos de tradição anglo-saxônica. Tais restrições surgem a partir da representação do mundo centrada na área de conhecimento, por meio de restrições lexicais e traços sintáticos característicos (p. 93). Novamente, temos uma descrição semelhante a dos estudos que adotam a Linguística de *Corpus* com base em Biber e se voltam para a caracterização dos discursos acadêmicos valendo-se dessas classificações.

A respeito da questão do contorno que a linguagem assume no texto especializado, uma outra acepção de análise é da ordem da relação entre texto especializado e gênero discursivo. Embora o texto especializado e os gêneros discursivos tenham trajetórias distintas nos estudos linguísticos e, aparentemente, possam parecer ter mais diferenças do que semelhanças, essas duas entidades comunicativas também apresentam encaixes. Embora, como ponto de partida, reforce-se aqui a asserção de Possamai (2004), quando diz que

Quando falamos em um gênero, estamos falando do geral, de um agrupamento de eventos comunicativos (entre esses o texto) que se assemelham. Por outro lado, um texto especializado é o particular, é a matéria que compõe os textos que pertencem um gênero (2004, p. 175).

De fato, quando nos referimos a gênero, tal como postulado por Bakhtin e Swales, temos diante de nós um fenômeno que se relaciona ao campo do discurso e que não está voltado diretamente para um tipo de linguagem específica. O texto especializado, por sua vez, além de se realizar por critérios textuais e discursivos, tais como as particularidades de sua macroestrutura, como as relações de coerência entre os seus elementos e a utilização de unidades sintáticas, lexicais, morfológicas e grafo-fonéticas, também se orienta por critérios temáticos. Afinal, está no seio da área científica ou técnica. Portanto, retomando agora as concepções de Swales e suas preocupações com o ensino e aprendizagem de ESP, temos aqui um aspecto em comum que é a comunicação científica. Ou seja, temos um critério também da esfera temática, se não específica, em um âmbito amplo.

Embora Swales (1990; 1992; 1998) não se aprofunde nas distinções lexicais dos artigos científicos das diferentes comunidades discursivas, como bem observa Zilio (2009), assim como mencionado anteriormente que ele tampouco se preocupa especificamente como item frequência de ocorrência de itens léxico-gramaticais. No entanto, como já notado, se nos apropriamos de gênero como um construto voltado para “variedades definidas situacionalmente” (BIBER, 1995, p.1) e agregamos a noção de *registro*, que inclui preocupações de caráter lexical e sintático, bem como questões de frequência, estamos em posição de observar trajetórias não apenas próximas, como bastante harmônicas.

Hoffmann (FINATTO e ZILIO, 2015), por outro lado, revela que o fato de o conceito de gênero textual ter sido definido de várias formas – e acrescente-se aqui, por várias perspectivas⁷⁰ – “não impediu que a pesquisa de linguagens especializadas realizasse uma observação mais precisa desses vários gêneros textuais que eram mais facilmente evidenciados na comunicação especializada”, (p. 100).

Acerca dos gêneros especializados, Hoffmann (*op. cit.*) postula que pertencem a “uma classe especial de gêneros textuais. A produção e recepção desses gêneros requer conhecimento especializado, além do conhecimento mundano, para as quais estão em vigor limitações mais rigorosas” (p. 101). Interessantemente, tanto Swales quanto Hoffmann “têm em mente um fim didático para a descrição dos gêneros” (KILIAN e FINATTO, 2015, p. 126).

As características estruturais (internas do texto) servem principalmente para descrever (gêneros de) textos especializados e, por sua ligação estreita com categorias e formadores linguísticos, desempenham um papel muito importante no ensino de línguas estrangeiras, sobretudo no desenvolvimento da chamada competência textual e na redação de textos autênticos. Já as características funcionais (externas ao texto) oferecem um potencial de explicação para sua escolha e uso conscientes. (HOFFMANN, 2015, p. 137.)

O interesse das abordagens de ensino pelo estudo do texto especializado nas últimas décadas representa um ponto vital para nossa pesquisa.

Como bem observa Kilian, “a didática mostrou a necessidade de uma reflexão teórica sobre o objeto dessa prática” (2007, p. 45). Os estudos em EAP, desse modo, de acordo com Hoffmann (2015, p. 41), podem se beneficiar didaticamente da comparação de fronteiras entre as linguagens especializadas e o ensino do léxico específico em função da seguinte tripartição:

1. determinados recursos linguísticos aparecem em todas as sublinguagens⁷¹ (o léxico geral da língua, por exemplo);
2. determinados recursos linguísticos aparecem em todas as linguagens

⁷⁰ A propósito, enquanto Swales, como notado anteriormente enfatizou o propósito comunicativo em sua definição, Hoffmann acentua a ação, ao apresentar gênero como “instrumento e resultado de um ato” (KILLIAN e FINATTO, 2015, p. 126.).

⁷¹ “Sistema parcial ou subsistema da linguagem que se atualiza em textos de âmbitos comunicativos específicos. Pode-se também dizer: uma sublinguagem é um recorte de elementos linguísticos e de suas relações estabelecidas em textos de uma temática delimitada.” (HOFFMANN, 2015, p. 40.) Hoffmann acrescenta que uma sublinguagem se traduz, geralmente, por parâmetros quantitativos, ou seja, pela frequência de determinadas manifestações linguísticas. Enquanto linguagens especializadas são sublinguagens, nem todas as sublinguagens são linguagens especializadas.

especializadas (o léxico científico, por exemplo);

3. determinados recursos linguísticos aparecem apenas em uma determinada linguagem especializada (o léxico da área da saúde, por exemplo).

Na verdade, “a especificidade das sublinguagens desempenham também um papel, na medida em que determinam a seleção e a organização do material didático-linguístico por parte do professor” (HOFFMANN, 2015, p. 40). Coerentemente, o pesquisador aponta que “o ensino ou a aquisição do vocabulário especializado não é o objetivo principal das aulas de línguas estrangeiras, mas, antes disso, deve ser uma parte da evolução da competência comunicativa”.

Por fim, podemos também pensar em como a problemática do texto especializado se relaciona à conceituação de especialização, que “permite quantificadores do tipo mais/menos/ algo/pouco/muito. Como determinar de maneira analítica que um texto é mais (ou menos) especializado que outro?” (CIAPUSCIO, 1998, p. 46). Para lidar com essa questão, Hoffmann (1998) postula dois eixos de classificação: um eixo horizontal que diz respeito a critérios temáticos e um eixo vertical que se relaciona ao grau de abstração da linguagem.

Krieger e Finatto (2004) esclarecem que o eixo horizontal contabiliza a variedade de temática de diferentes disciplinas, atividades e profissões existentes, enquanto o vertical compreende níveis de abstração.

Esses níveis levam a reconhecer os graus de especialidade nas comunidades profissionais, conforme o uso mais ou menos frequente, tanto de determinadas estruturas sintáticas quanto de termos técnicos. Por exemplo, as comunidades científicas entre especialistas pertencem ao nível mais elevado de abstração e caracterizam-se pela utilização de determinadas formas de expressão linguísticas, além de símbolos artificiais. (p.114)

Aqui também são levados em consideração o contexto e os parceiros, ou os cientistas como protagonistas dessa relação dialógica⁷². A partir desses critérios, Hoffmann

⁷² Cabré (1999), assim com mencionado acima a respeito das proposições de Hoffmann, defende que só podem ser “produtores de comunicações técnico-científicas os profissionais de nível restrito, aqueles que têm um conhecimento específico da temática, conhecimento adquirido previamente pela aprendizagem acadêmica ou profissional” (1999, p. 153). Com relação a temáticas, no entanto, são consideradas especializadas aquelas que se caracterizam como objeto explícito de ensino acadêmico ou profissional.

(1998) estabelece uma ordem para as linguagens especializadas que começa nas ciências básicas, passando pelas ciências experimentais, aplicadas e técnicas, seguindo-se a linguagem da produção material e finalizando com a do consumo. A linguagem especializada para esse autor, não se trata, por sua vez, de um todo monolítico e homogêneo, mas um complexo de tipologias distintas, que representam fundamentos para a compreensão do fenômeno da comunicação especializada.

Já Ciapuscio (1998; 2003) introduz a importante noção de variação conceitual ao postular níveis para os textos especializados. Tais níveis são estabelecidos como funcionais ou situacionais e formais em uma relação direta com os graus distintos de especialização, que caracterizam determinadas classes textuais. Nas palavras da autora

[...] os fatores de índole funcional e situacional (interlocutores, classe textual, âmbito discursivo) condicionam a seleção, o tratamento e os limites da variação formal e conceitual da terminologia. Por outro lado, pode afirmar-se que o ângulo terminológico oferece argumentos de peso para a tipificação das classes de textos (*abstract*, notícia de divulgação científica) no nível de formulação (cf. classes textuais) e para determinação de seu grau de especialidade, a partir de fundamentos linguísticos (CIAPUSCIO, 1998, p. 63).

Podemos reconhecer essas asserções na constatação de que “mesmo abordando um mesmo tema científico, geralmente, as terminologias utilizadas não coincidem” (KRIGER & FINATTO, 2004, p. 117). Isso acontece em função de haver muitas variações denominativas relativas ao maior ou menor grau de especialização de um texto, ou seja, uma notícia de divulgação científica publicada em uma revista não especializada terá terminologias distintas de sua publicação original. Variará também seu grau de detalhamento ou densidade informativa, se os destinatários forem leigos ou especialistas, aspecto que demonstraremos em nossa coleção de dados.

Ao longo deste capítulo, buscou-se fornecer elementos para permitir ao leitor entender a “totalidade funcional e estrutural” (HOFFMANN, 2004, p. 79) constituída pelo texto especializado. A intenção, para além de situar o texto especializado por uma perspectiva diacrônica, também foi de apresentá-lo como alavanca para participação no mundo da ciência e, por extensão, para construir e modificar a representação do fazer científico.

3. CAPÍTULO 3

“I would have written a shorter letter, but I did not have the time”

Blaise Pascal, 1657.

“Brevity is the soul of wit.” (Hamlet Act 2, scene 2, 86–92)

William Shakespeare

ABSTRACTS, UMA VIRADA NA HISTÓRIA DOS TEXTOS ACADÊMICOS

Neste capítulo, aborda-se a constituição específica do gênero textual sobre o qual a investigação se debruça. Para tanto, percorremos a trajetória da criação do *abstract* no âmbito da escrita em Ciências da Saúde, abordando a ampliação de seus objetivos e as motivações de pesquisa no tema. Do mesmo modo, estudamos a caracterização dessa classe de texto a partir de seus principais pesquisadores, com base na tradição de pesquisa anglo-saxônica.

3.1. Das origens e dos objetivos

A expansão da ciência na segunda metade do século XX determinou o surgimento de uma gama de publicações especializadas em diferentes esferas do conhecimento, principalmente na área biomédica. Cilvette e Pérez (2006) apontam que é, em larga medida, pela circulação dessas publicações, as quais representam a principal fonte de atualização dos profissionais do âmbito clínico, que o campo avança.

A multiplicação da oferta de periódicos especializados culminaria, não apenas na ampliação da disponibilidade de *know-how* aplicado disponível, como, também, em limitações à capacidade de especialistas e pesquisadores para acompanhar integralmente os avanços da ciência. A restrição, por conta da abundância de textos publicados, impulsionou a urgência de a comunidade científica criar um modo de dar acesso sintetizado ao conteúdo dos estudos, uma forma de apresentar a essência das pesquisas com poucas palavras.

É nesse contexto que emerge um novo tipo de escrito científico que se torna primordial para facilitar o acesso e a seleção de produções acadêmicas pelo leitor, ao conjugar o máximo de informação a um mínimo de palavras. Originário do latim ‘*abstractus*’, que significa *extrair de* ou *separar*, o termo *abstract* define a síntese de um texto acadêmico, seja de um artigo, de uma tese ou dissertação, que começa a ser incorporado no final dos anos 60 e que, a partir dos anos 70, faz-se presente em todas as publicações de renome da área da saúde (SALAGER-MEYER, 1991).

Após os objetivos para o novo gênero serem reconhecidos pela comunidade discursiva que lhe deu origem, nos termos de John Swales (1990), que foi um dos precursores nesses estudos, a razão de ser do texto determinou a estrutura do discurso e delimitou as escolhas de conteúdo e estilo. Ou seja, “a produção exemplar de importantes especialistas, também conduziu à elaboração de planos de construção textual relativamente fixos” para o manuscrito (HOFFMANN, 2015, p. 28). Da mesma maneira, novas finalidades para sua adoção foram surgindo à medida que o uso se consolidava. Esses propósitos foram contribuindo para a consolidação da relevância e independência do *abstract* no cenário da divulgação das pesquisas científicas.

Primeiramente, a partir da proposição de duas revistas especializadas de expressão, o *Journal of the American Medical Association*⁷³ e o *Canadian Medical Association Journal*⁷⁴, convencionou-se que o *abstract* seria incluído no início do texto da pesquisa, o que fortaleceu a função de ponto de entrada para os textos acadêmicos. A medida foi de enorme validade para os sistema de busca eletrônica.

A evolução tecnológica produziria uma proliferação dos sistemas de informação eletrônica, especialmente a partir do advento da *internet*. Consequentemente, os usuários dos textos acadêmicos precisavam lidar com formas de filtrar a informação. Os profissionais do campo, tais quais bibliotecários, indexadores e especialistas em desenvolvimento de sistemas, passariam então a aceitar o *abstract* mais amplamente a partir dos anos 70, como ferramenta de recuperação de informação (CHAN e FOO, 2004).

Por meio de texto conciso como o do *abstract*, o estudo completo poderia ser facilmente recuperado. Por esse motivo, os profissionais da informação estabeleceriam parâmetros para a organização do texto. Os mais significativos foram estabelecidos pelo *American National Standard Institute* (ANSI, 1979)⁷⁵ e o *International Standards Organization* (ISO,1976)⁷⁶.

A partir de então, os motores de busca de bancos de dados de literatura acadêmica passaram a indexar primeiramente e, em alguns casos, apenas o *abstract*, no lugar do texto integral de um estudo. Essa autonomia destacaria a função de entidade independente do gênero. Com isso, igualmente, o novo manuscrito acadêmico assumiria o papel de auxiliar os leitores na decisão de ler ou não o texto completo. Do mesmo modo, o *abstract* determinaria a reimpressão ou aquisição da versão eletrônica da investigação, já que, em

⁷³ <http://jama.jamanetwork.com/journal.aspx> (Último acesso em 01/11/2015).

⁷⁴ <http://www.cmaj.ca> (Último acesso em 01/11/2015.)

⁷⁵ <http://www.ansi.org/Accreditation/product-certification/Default.aspx> (Último acesso em 15/11/2015.)

⁷⁶ <http://www.iso.org/iso/home.html> (Último acesso em 15/11/2015).

função da questão de direitos autorais, o acesso ao conhecimento veiculado através dos periódicos especializados, muitas vezes, não é democratizado⁷⁷.

Até a metade dos anos 80, o *abstract* era apresentado em formato desestruturado, em estilo de texto corrido, com a extensão de um parágrafo. Embora o padrão estrutural IMRD (Introdução, Metodologia, Resultados e Discussão) tivesse sido proposto inicialmente por editores de periódicos especializados da área médica na década de 40, (SALAGER-MEYER, 1990), os textos dos *abstracts*, frequentemente, não seguiam o padrão especificado e eram, até mesmo, incompletos.

O seguinte texto representa um *abstract* desestruturado:

Quadro (1): *Abstract* Estruturado

Is Everybody Happy? (JAMA. 1969;207(5):942. doi:10.1001/jama.1969.03150180072014.)

*The scientists, it would seem, have finally got on to something really big, something which has baffled philosophers for centuries. Whereas we have been brought up on the old saw that money won't buy happiness, a study conducted by the US Public Health Service would seem to indicate the contrary. The findings were substantiated by a survey of a random sample of some 1,400 Puerto Ricans who were interviewed on the island during a six-month period in 1963-1964. The sample was validated by questioning 114 outpatients of a psychiatric hospital, also in Puerto Rico. Probably the most difficult part of the study for the investigators was finding a definition of the term "happiness." Webster's definition would seem to be quite apt for the nonscientist, but it hardly lends itself to statistical analysis, which is what science seems to be all about these days.*⁷⁸

Fonte: Freitas (2016)

Em 1987, o Comitê Internacional de Editores de Jornais Médicos e o *Ad Hoc Working Group for Critical Appraisal of Medical Literature* propõem a criação do *abstract* estruturado para artigos acadêmicos, em que cada parte da pesquisa é explicitamente nomeada em formato de itens, como no texto que se segue:

Quadro (2): *Abstract* Estruturado

A Whole-Grain-Rich Diet Reduces Urinary Excretion of Markers of Protein Catabolism and Gut Microbiota Metabolism in Healthy Men after One Week.

Introduction: *Epidemiological studies consistently find that diets rich in whole-grain (WG) cereals lead to decreased risk of disease compared with refined grain (RG)-based diets. Aside from a greater amount of fiber and micronutrients, possible mechanisms for why WGs may be beneficial for health remain speculative.*
Purpose: *In an exploratory, randomized, researcher-blinded, crossover trial, we measured metabolic profile differences between healthy participants eating a diet based on WGs compared with a diet based on RGs.*
Methodology: *Seventeen healthy adult participants (11 female, 6 male) consumed a controlled diet based on either WG-rich or RG-rich foods for 2 wk, followed by the other diet after a 5-wk washout period. Both diets were the same except for the use of WG (150 g/d) or RG foods. The metabolic profiles of plasma, urine, and*

⁷⁷ Correntemente, parte dos artigos na literatura biomédica está disponível através da MEDLINE, sistema de busca bibliográfica interativa, acessível através do banco de dados PubMed (Disponível em <http://www.ncbi.nlm.nih.gov/pubmed/> . Último acesso em 31/10/15.)

⁷⁸ Disponível em: <http://jama.jamanetwork.com/article.aspx?articleid=343532> (Último acesso em 14/11/2015.).

fecal water were measured using ^1H -nuclear magnetic resonance spectroscopy and gas chromatography-mass spectrometry (plasma only).

Results: After 1 wk of intervention, the WG diet led to decreases in urinary excretion of metabolites related to protein catabolism (urea, methylguanadine), lipid (carnitine and acylcarnitines) and gut microbial (4-hydroxyphenylacetate, trimethylacetate, dimethylacetate) metabolism in men compared with the same time point during the RG intervention. There were no differences between the interventions after 2 wk. Urinary urea, carnitine, and acylcarnitine were lower at wk 1 of the WG intervention relative to the RG intervention in all participants. Fecal water short-chain fatty acids acetate and butyrate were relatively greater after the WG diet compared to the RG diet.

Implications: Although based on a small population and for a short time period, these observations suggest that a WG diet may affect protein metabolism.⁷⁹

Fonte: Freitas (2016)

Essa classe de *abstract* passa a ser adotada cada vez mais frequentemente no campo biomédico e contribui para tornar o manuscrito mais informativo. Alguns pesquisadores, a exemplo de Cilvette e Pérez (2006), avaliam esse caráter informativo do *abstract* estruturado como fundamental, uma vez que o processo de tomada de decisões clínicas é, muitas vezes, pautado pelo *abstract*, como entidade autônoma, nos casos de pesquisas que relatam investigações novas ou casos clínicos (p. 64-65).

A partir do final da década de 80, o gênero começa a funcionar como um mapa de leitura do texto que introduz (HUCKIN, 2001). Ele também é alçado à função de representar a pesquisa completa, como base para a seleção de trabalhos submetidos para apresentação em conferências acadêmicas e para publicações em periódicos especializados (SWALES e FEAK, 2009). Nesse período, além dos objetivos apontados, agrega-se o papel de facilitador do processo de revisão pelos pares (AD HOC GROUP, 1987). Pesquisadores do campo biomédico, passam a defini-lo como “janela de trabalho disponível para uma avaliação acessível”⁸⁰ (MCINTOSH, 1993, p. 3).

Por fim, o fato de grande parte dos estudos acadêmicos do século XX terem sido desenvolvidos no hemisfério norte, portanto, em língua inglesa, gerou uma expectativa de que as investigações conduzidas no resto do mundo pudessem circular pelos grandes centros de pesquisa. Ademais, o fator de impacto estabelecido como medida de qualidade para as publicações especializadas determina que as publicações recebam citações como medida da circulação atingida. Por essa razão, o *abstract* passa a ser quase sempre apresentado em língua inglesa, mesmo quando relacionado a um texto produzido em outra língua (WOOD, 2001). Essa particularidade acaba por alargar a visibilidade do manuscrito por conta da redação na língua franca da ciência (FLOWERDEW, 2002). Ressalta-se que essa tradição se estendeu

⁷⁹ Disponível em: <http://jn.nutrition.org/content/143/6/766.abstract#fn-1> (Último acesso em 14/11/2015.)

⁸⁰ “The window of work available for easy evaluation” (MCINTOSH, 1997, p. 3).

para além dos artigos acadêmicos, abrangendo também as teses e dissertações, dentre os principais manuscritos de pesquisa.

Correntemente, o *abstract* detém papel consolidado no campo da saúde, bem como em muitas outras áreas do conhecimento acadêmico. Uma prova da vitalidade alcançada pelo manuscrito é o fato de os estudiosos detectarem nesse diferentes estruturas de texto, relativas às diferentes áreas do saber que o empregam (BIASI-RODRIGUES *et al.*, 2009, SHEHZAD, 2010).

Se comparamos essa história de poucas décadas à trajetória do artigo acadêmico biomédico, concebido primeiramente em estilo de carta, nos idos de 1665 (GUIMARÃES, 2006; SOLACI e PEREIRA, 2004), estamos diante de uma rota ainda muito breve. Não obstante, o novo manuscrito se estabelece como gênero em um contexto de intensa ebulição científica, traduzida por uma inflação de periódicos especializados, quando passa a ser valor dizer menos para atingir mais. Nos termos de Hyland,

O artigo acadêmico representa, em sua essência, a codificação do conhecimento disciplinar, em que os autores buscam persuadir as comunidades em que se inserem a aceitar suas asserções como formas de conhecimento legítimo e reconhecido. Os *abstracts* [...] detêm uma função tanto mais modesta quanto urgente: persuadir os leitores de que vale a pena ler o artigo. Figuram, portanto, como uma representação seletiva antes de uma tentativa de fornecer ao leitor o conhecimento detalhado do conteúdo de um artigo⁸¹. (HYLAND, 2000, p. 64.)

A conquista do *status* de manuscrito independente corrobora para fortalecer o protagonismo dessa *representação seletiva*. Afinal, o *abstract* se tornou, em boa medida, responsável pela leitura ou inclusão/exclusão de uma investigação em um periódico ou conferência. Portanto, compete-lhe promover acesso à representação integral do conhecimento disciplinar, referida por Hyland (*op. cit.*) acima. Tarefa que, por conseguinte, não parece modesta.

De acordo com Graetz (1985), uma das precursoras nos estudos do tema, a maior parte dos leitores não chega a ler o texto do estudo completo (p. 23). Para Pinto e Lancaster (1991), os textos completos das pesquisas geralmente contêm detalhes que dificultam a identificação de conceitos significativos, enquanto o *abstract* consegue filtrar

⁸¹ “The research article is, in essence, a codification of disciplinary knowledge, where writers seek to persuade their communities to accept their claims and certify them as recognized and legitimate knowledge. Abstracts [...] have both a more modest and more urgent purpose: to persuade readers that the article is worth reading. It is therefore a selective representation rather than an attempt to give the reader exact knowledge of an article’s content.” (Hyland, 2000, p. 64)

detalhes desnecessários. Ou seja, nesses casos, recai exclusivamente sobre o gênero a função de disseminar o conhecimento científico.

Com base nesses méritos, o *abstract* pode ser acolhido como uma virada na história dos textos acadêmicos. Portanto, a compreensão e redação adequadas do gênero são fundamentais para a pesquisa que sintetiza.

3.2. Das caracterizações e do posicionamento das investigações

De acordo com Hoffmann (2015), “a caracterização de alguns gêneros textuais, chega quase a uma tentativa de definição, como é o caso do *abstract*” (p. 142). Na acepção de Gläser (1990), “o *resumo*⁸² publicado em revistas especializadas internacionais é um texto curto derivado de um trabalho científico original, que se caracteriza pela densidade de conteúdo e um alto grau informacional” (p. 126). Ao enfatizar os aspectos *brevidade*, *densidade de conteúdo* e *alto grau informacional*, essa descrição reflete a complexidade do gênero.

Swales (1990) considera o processo de redigi-lo um “rito de passagem” que se concretiza por uma demonstração de domínio da linguagem e da retórica especializadas, para ingresso na comunidade científica. Enquanto Orsan (2001) avalia que

a redação de um *abstract* não é tarefa trivial, uma vez que o trabalho não permite redundâncias e impõe aos autores a adoção de muitos itens lexicais compostos. Como Halliday (1993) aponta, em textos científicos, incluindo *abstracts*, a densidade lexical é alta, o que torna difícil tanto ler quanto escrever esses textos (ORSAN, 2001, p. 334).⁸³

As questões apontadas pelos autores acima sintetizam muito da motivação deste trabalho em explorar o gênero, assim como as motivações de tantos outros investigadores que se voltaram para o tema. Até mesmo porque a produção de um *abstract* fluente é tarefa complexa para pesquisadores experientes e muito mais para aqueles que não transitam fluentemente pela língua do texto a ser redigido. A fluência, nesse último caso, está

⁸² Grifo da autora.

⁸³ “This is true, writing an abstract is not a trivial task given that it does not allow redundancies and forces the writers to use a lot of compound words. As Halliday (1993) points out, in scientific texts, which include scientific abstracts, lexical density is very high, which makes it difficult to both read and write such texts.” (ORSAN, 2001, p. 334)

relacionada tanto a aspectos discursivos de caráter amplo, da conta da funcionalidade dos textos, quanto às escolhas léxico-gramaticais (BIBER *et al.*, 1999). Assim, neste caso, o interesse é por explorá-lo a partir dos *itens lexicais compostos* a que Orasan (*op. cit.*) alude acima. Em sequência, averiguamos como essas preocupações vêm se refletindo nos interesses de outros pesquisadores.

Nesta parte da análise, caracteriza-se o gênero textual, com base em algumas das principais investigações do tema. Ao fazê-lo, busca-se situar tais investigações em relação aos posicionamentos epistêmicos que as norteiam, os quais vão estabelecendo concepções para explicar o objeto de seu interesse, apesar de o tema não ter sido tão extensivamente abordado quanto os artigos acadêmicos.

Dentre as principais perspectivas investigativas a partir das quais os *abstracts* de textos científicos têm sido abordados, incluem-se as análises de registro (SWALES, 1985; BIBER *et al.*, 2000) e as investigações centradas em gêneros acadêmicos (SWALES, 1990; HYLAND, 2004; SWALES e FEAK, 2009).

Por análises de registro, entende-se que um tipo de texto é definido a partir de padrões léxico-gramaticais (HALLIDAY E HASSAN, 1978; SWALES, 1985; BIBER *et al.*; 1999; BIBER, *et al.*, 2000; BIBER e CONRAD, 2009), conforme abordado anteriormente. Uma asserção fundamental dessa abordagem é de que “aspectos de contexto e características linguísticas podem ser alinhados” (HUNSTON, 1995, p.13). Não obstante, foram os postulados de Biber e Finegan (1986) que deram início às análises de registro computadorizadas, um método que identifica a variação nos textos pela ocorrência de características linguísticas, referido acima como impulsionador de muitos estudos do tema em Linguística de *Corpus* (LC).

Já os estudos de *gêneros acadêmicos* com vistas à aprendizagem de língua ganharam força na década de 1990, em boa medida, graças às pesquisas de John Swales. Retomando pontos já discutidos, esse autor assume que são os propósitos profissionais (SWALES, 1981; 1990) que definem os tipos de texto, cujas características linguísticas variam de acordo com os diferentes objetivos comunicativos dos membros de uma comunidade. Tais investigações tiveram por metas principais identificar e acessar as ações retóricas tomadas pelo escritor durante a produção textual, traduzidas em escolhas linguístico-discursivas específicas.

É na interseção das perspectivas mencionadas, quais sejam, as análises de registro com esteio em LC e as investigações centradas em gêneros acadêmicos, que este estudo se posiciona.

Cumpra mencionar também que o interesse dos profissionais da Ciência da Informação, da Computação e do Processamento de Línguas Naturais (PLN) por estabelecerem os parâmetros mencionados na seção anterior - ANSI, 1979 e ISO, 1976 - igualmente motivou estudos no tema. Esses pesquisadores buscaram organizar guias para essa classe de texto, de modo a apresentarem maior brevidade, acurácia e clareza em atender as necessidades dos usuários (LANCASTER, 1991; ORASAN, 2001). A extração automática de *abstracts* tem influência dos trabalhos de Van Dijk e Kintsch (1993) sobre gramática textual proposicional e dos estudos de Beaugrande e Dressler (1981), sobre resumos formados a partir da relação entre partes de textos.

No Brasil, destacam-se os trabalhos de Aluísio e Oliveira (1995), Aluísio e Ganteinbein (1997), Aluísio *et al.* (2001), Feltrim (2004) e Genovês *et al.* (2007), voltados para a criação de ferramentas de escrita acadêmica para a escritura de *abstracts*. No âmbito internacional, reconhecem-se, especialmente, os esforços de Narita *et al.* (2003) e Anthony (2006) para facilitar a redação do texto.

3.2.1. Do legado dos primeiros estudos

As investigações iniciais das décadas de 80 e 90 sobre o tema assumiriam tanto um caráter léxico-gramatical, quanto um caráter descritivo a partir da concepção de gêneros acadêmicos. A motivação das primeiras foi o ensino da língua inglesa para promover a leitura e a escrita em áreas aplicadas do conhecimento científico. As segundas se voltaram para apresentar planos estruturais para a organização do texto, tomando por base as motivações das diferentes comunidades especializadas que o adotam.

Dentre precursores dessas pesquisas encontram-se Van Dijk (1980), Swales (1981⁸⁴, 1984, 1985, 1990), Graetz (1985) e Myers (1988, 1990), Cremmins (1982) e Endres-Niggemeyer (1985, 1990, 1991).

⁸⁴ Note-se que, na verdade, John Swales lançou a monografia *Aspects of Article Introductions* sobre introduções de artigos acadêmicos em 1981, antes de lançar sua descrição para os gêneros acadêmicos (1990). Esse trabalho seria publicado, à época, quase que de forma marginal, *underground*, nos próprios termos de Swales (2011, p. 1). O estudo teve formato de um manuscrito em espiral com 120 cópias que, em função de algumas reproduções adicionais, chegou a cerca de 400 cópias em 1984, para ser relançado apenas em 2011. Foi a partir da obra lançada em 1990 que o trabalho desse linguista no tema ficaria amplamente conhecido.

Destacam-se, igualmente, nesta fase, as investigações de cunho sociológico de Bhatia (1993) e de Bazerman (1988). Ambos perceberiam o *abstract* e o artigo acadêmico como gêneros inter-relacionados, pertencentes à mesma colônia (BHATIA, *op. cit.*) ou ao mesmo sistema de gêneros acadêmicos (BAZERMAN, *op. cit.*).

Em estudos a partir de campos acadêmicos variados, Swales (1981,1984), Cremmins (1982), Nwogu (1989), Salager-Meyer (1990, 1992), Posteguillo (1996), Santos (1996) e Ventola (1997) rastrearão a presença e as formas de realização da estrutura IMRD, que havia sido sugerida anteriormente por editores de publicações da área biomédica.

Dentre esses pesquisadores, Salager-Meyer (1990) se voltou para o campo da Medicina e constatou que muitos *abstracts* em suas análises não só não eram fidedignos em relação aos textos que sintetizavam, como eram pouco informativos e apresentavam uma estrutura ilógica. Dos 77 abstracts analisados, apenas 52% seguiam o formato padrão IMRD. Esse estudo tem, desde então, tido bastante repercussão dentre os pesquisadores de EAP com foco em saúde.

Graetz (1985) igualmente buscou o padrão IMRD em *abstracts* do *American Journal of Public Health*⁸⁵, a partir de uma preocupação de potencializar competência de leitura no segmento acadêmico. Ao desenvolver esses estudos, a investigadora chegaria a uma tipologia do gênero, com a seguinte classificação:

- a. *Abstract* Informativo – de natureza descritiva, deve apresentar ao leitor as principais características do estudo - problema, metodologia, resultados e conclusão - na mesma sequência argumentativa do texto maior, para possibilitar a compreensão desse;⁸⁶
- b. *Abstract* Indicativo – de natureza descritiva, deve apresentar o estudo em poucas palavras, como um sumário narrativo em que se inclui sucintamente os conteúdos – geralmente problema, metodologia e escopo - para promover uma compreensão da natureza geral e da finalidade da pesquisa sem, no entanto, abordar o passo a passo da investigação;

⁸⁵ <http://ajph.aphapublications.org> (Último acesso em 14/11/2015.)

⁸⁶ A Associação Brasileira de Normas Técnicas (ABNT) também estabelece que o *abstract* informativo ou analítico é capaz de dar conta dessa representação ao introduzir os propósitos, metodologia, resultados e conclusões do trabalho. Ou seja, cumpre a função de reduzir o texto em tamanho, mantendo-se as ideias principais. Em extensão, atinge de 150 a 500 palavras para trabalhos acadêmicos, teses dissertações e relatórios técnico-científicos, assim como de 100 a 250 palavras para os artigos e periódicos (ABNT, 2015). Essa extensão é variável com base nos critérios de submissão das diferentes publicações especializadas. <http://www.abnt.org.br/normalizacao/lista-de-publicacoes/abnt> (Último acesso em 05/11/2015.)

c. Abstract Crítico – de cunho descritivo e crítico, é de uso mais raro e deve agregar uma avaliação do conteúdo da pesquisa, expressa no posicionamento do autor, sendo extenso a ponto de dispensar a leitura do texto completo para compreensão dos argumentos da pesquisa.

Estabelecidas as distinções, a autora constatou que, enquanto a inclusão já era abundante na área da saúde, poucas pesquisas do campo das humanidades continham *abstracts* àquela época. Por isso, voltou-se para investigações biomédicas, tendo eleito os textos de classificação informativa como foco de interesse.⁸⁷

Feita a classificação dos textos, Graetz (1985) registraria as características lexicais então. A estudiosa notaria uma tendência ao uso do passado (o presente foi identificado mais em comentários do que em narrativa)⁸⁸, de terceira pessoa e voz passiva, da adoção de mais locuções do que orações, assim como de mais palavras do que de locuções. Por vezes, há o uso de verbos de ação. A autora constataria, ademais, a ausência de negativas, orações subordinadas, abreviações, jargões, símbolos, repetições, superlativos, adjetivos, ilustrações, notas de rodapé e exemplos. Ela então apresentou uma descrição para o *abstract* informativo ideal:

Deve ser uma narrativa contínua, produzida em sentenças completas. Não deve adotar parágrafos separados para a caracterização de problema, sumário, introdução, metodologia, etc. Deve refletir a organização do artigo, seguindo a ordem exata do texto, se possível. Deve fornecer mais informação do que o apresentado no título. Esse último não deve ser repetido. O abstract deve ser breve, não desperdiçar palavras, ainda assim, deve ser longo o suficiente para expressar as concepções do autor. Não deve apresentar crítica ou viés; pois não se trata de uma revisão. Deve evitar ambiguidades, ser inteligível, legível e apresentar-se como um texto independente. Tanto quanto possível, deve ser redigido nos próprios termos do autor.⁸⁹ (GRAETZ, 1985, p. 23.)

⁸⁷ Do mesmo modo como essa autora, concentramos nossa atenção nos textos do tipo informativo, porque defendo a noção de que um *abstract* representativo de um texto acadêmico dispensa a leitura do texto original para o conhecimento geral do assunto abordado, sem, contudo, tecer juízo de valor a respeito da pesquisa em questão.

⁸⁸ Tendência confirmada por Salager-Meyer (1992) em um estudo de *corpus* com uma coleção de 84 *abstracts*, em que a autora constata que, em mais da metade dos textos, mantém-se uma prevalência de uso do passado.

⁸⁹ “*It should be a continuous narrative, written in whole sentences. It should not use separate paragraphs for the commonly recurring features of problem, summary, introduction, method, etc. It should reflect the organization of the article, by following the exact order if/possible. It should provide more information than the title. The title should not be repeated. The abstract should be brief, not waste words, yet be long enough to convey the author's concept. It should be non-critical, unbiased; it is not a review. It should be unambiguous, intelligible, readable and a complete item in its own right. It should be written in the author's own language (as far as possible)*”. (GRAETZ, 1985, p. 23.)

Essa caracterização ainda se mostra bastante afinada com os *abstracts* das publicações das revistas especializadas internacionais da coleção de dados deste estudo. Ou seja, a definição remete à prática do que a comunidade científica continua estabelecendo como norma de circulação para os usuários da língua inglesa.

A respeito da extensão sugerida, porém não especificada por Graetz (1995), Swales e Feak (2009) asseveram que a maioria dos jornais especializados demanda que os *abstracts* tenham entre 150 e 200 palavras quando de padrão desestruturado, em oposição aos estruturados, classe sobre a qual recai a maior parte dos textos de nosso *corpus*. Esses últimos podem ser mais longos⁹⁰. Quanto às teses, os dois autores mencionam que esses *abstracts* costumam atingir até 350 palavras; enquanto os que são submetidos a conferências são os mais longos, podendo chegar a 500 ou mais palavras.

Graetz (1985) ainda destacaria que

Embora a função primária dos *abstracts* seja de indicar e prever a estrutura e o conteúdo do texto que o segue, funcionam como *tipo independente de discurso*⁹¹ também. Em função de serem discursos em si, espera-se que sejam igualmente coesivos e coerentes.⁹² (GRAETZ, 1985, p.3)

Essa asserção sintetiza aspectos fulcrais acerca de um objeto de estudo que atingiria amplitude na década seguinte, através da divulgação das pesquisas de autoria de Swales.

Em uma publicação 1993, Swales concordaria com Graetz (*op. cit.*) no que diz respeito aos *abstracts* funcionarem como um tipo independente de discurso⁹³. Não obstante, ao darem sequência a esses estudos, Swales e Feak (2009) estabeleceriam distinção entre gênero e sub-gênero ou gênero parcial.

A partir da concepção de gênero como texto organizado para atingir uma gama de propósitos comunicativos e em função de conceberem o *abstract* como uma seção de um texto maior, Swales e Feak (*op. cit.*) descrevem-no como gênero parcial. Outrossim,

⁹⁰ Os periódicos da coleção internacional de dados desta pesquisa estabelecem as seguintes extensões: *British Journal of Pharmacology*: máximo de 250 palavras; *American Journal of Clinical Nutrition*: não mais de 300 palavras; *Annals of Internal Medicine*: exatamente 275 palavras.

⁹¹ Grifo meu.

⁹² “Although the primary function of abstracts is to indicate and predict the structure and content of the text which follow them, they function as an independent type of discourse as well. Because they are discourses in their own right they are expected to be both cohesive and coherent.”

⁹³ “Abstracts function as independent discourses as well as being advance indicators of the content and structure of the following text.” (Swales, 1993, p.179.)

defende-se aqui que o mesmo se sustenta como um gênero em si, uma vez que é independente a ponto de encontrarmos publicações apenas compostas de *abstracts*, tais quais anais de apresentações em eventos acadêmicos. Gläser (1991) também ratifica esse posicionamento. Gledhill (1995) refere-se ao *abstract* como gênero dotado de “inesperada complexidade, que frequentemente é considerado um *produto curto e seco*⁹⁴ (ou pior um excerto), derivado de um texto original”⁹⁵ (p. 14). Alinho-me, do mesmo modo, a Cilvette e Pérez (2006), quando essas autoras alegam que é questionável afirmar que o *abstract* é parte de um texto maior, já que, na tradição de pesquisa anglo-saxônica, certas áreas de conhecimento acadêmico não os incluem. Podemos citar como exemplos a música, literatura e a filosofia. Assim as duas autoras também apregoam que, em relação aos artigos de pesquisa, os *abstracts* podem perfeitamente ser aceitos como gênero textual em função de compartilharem de um *status* independente, bem como por apresentarem seus próprios propósitos comunicativos (p. 64). Em defesa desse posicionamento, encontram-se os entendimentos de Biber and Conrad (2009). Esses autores descrevem as seções dos artigos acadêmicos (*abstract* incluso), como gênero em si, apenas embebido em um gênero maior (p. 33). Ken Hyland (2004) também argumenta que, enquanto os *abstracts* de fato compartilham de características da pesquisa completa, “diferem na função, na estrutura retórica e na nas realizações linguísticas” (p. 74)⁹⁶. Sustenta-se aqui que esses últimos argumentos embasam estruturalmente a realização do texto como um gênero em si.⁹⁷

Retomemos agora a citação de Graetz (1985) da página anterior no que diz respeito às expectativas de coerência e coesão em relação ao *abstract*, pelo fato de o mesmo representar um discurso em si. Pela perspectiva da Teoria da Estrutura Retórica (MANN e THOMPSON, 1988), referida anteriormente no capítulo 1, um texto se estrutura de modo *coerente* em função de existir um objetivo para suas partes, alguma razão plausível para sua inclusão, que é evidente aos olhos de um leitor experiente. Há, desse modo, uma função desempenhada por cada segmento que compõe o texto, pela perspectiva da organização de um manuscrito. O segredo da coesão textual, por sua vez, está em fazer a costura característica para justapor as partes internas que compõem o todo.

Para Swales (1990; 1994), essas características se organizam a partir de modelos que dão forma à estrutura do texto e servem para orientar os membros especialistas

⁹⁴ Grifo meu.

⁹⁵ “*unexpected complexity of a genre that is often considered to be a cut and dry product (or worse, an extract) derived from an original text*” (GLEDHILL, 1995, p. 11).

⁹⁶ “... *differ in their function, in their rhetorical structure and in their linguistic realizations*”.

⁹⁷ Agradeço à Profa. MA. Cristina Gibk, do Curso de Letras da UNISINOS, pela calorosa discussão intelectual em fortalecimento a meu posicionamento.

das comunidades discursivas quanto às escolhas de conteúdo e de estilo. Ao mesmo tempo que os modelos norteiam os membros, são delimitados por suas motivações, em termos de formatação esquemática do manuscrito. Quando Swales (1990) lançou os critérios de definição de gênero acadêmico, ele também estabeleceu uma descrição organizacional das convenções para introdução de artigos acadêmicos,⁹⁸ que se tornaria muito difundida. Para ele, esse modelo se estabelece a partir de um formato padrão, não no sentido limitante, mas em um sentido organizacional.

A estrutura proposta ficou conhecida como modelo *CARS*⁹⁹ e compreende a descrição dos segmentos que desempenham funções específicas no texto, denominados movimentos retóricos. Swales (2004) estabelece que são esses segmentos que se encarregam de executar as funções comunicativas coerentes nos textos (p. 228).

Variada nomenclatura tem sido utilizada para rotular as unidades de informação observadas a partir desse formato: *moves* e *steps* (SWALES, 1990), *moves* e *submoves* (SANTOS, 1999), *moves* e *subfunctions* (MOTTA-ROTH, 1995), *moves* e *strategies* (ARAÚJO, 1999) e unidades retóricas (MEURER, 1997).

O modelo *CARS* está dividido em três movimentos¹⁰⁰ que abrangem passos específicos:

1. Movimento 1 – Estabelecer o Território

Passo 1: Estabelecer a importância da pesquisa e/ou

Passo 2: Fazer generalizações quanto ao tópico e/ou

Passo 3: Revisar a literatura

2. Movimento 2 – Estabelecer o Nicho

Passo 1a: Contra-argumentar ou

Passo 1b: Indicar lacuna(s) no conhecimento já estabelecido ou

Passo 1c: Provocar questionamentos ou

Passo 1d: Continuar a tradição

⁹⁸ Note-se que a referência à *introdução de artigos acadêmicos* como forma de aludir à *abstract* é polissêmica, pois *introdução*, nos termos de Swales e Feak (2009), remete à seção específica do texto acadêmico que sucede ao *abstract*.

⁹⁹ *Create-A-Research-Space*.

¹⁰⁰ De acordo com Motta-Roth (1995), movimentos são “unidades de informação com extensão de uma ou mais sentenças, as quais realizam uma função em particular e, junto com outros ‘*moves*’, constitui a estrutura informacional total presente no texto” (Disponível em: <http://www.ufsm.br/labler/publi/anlise.htm> Último acesso em 15/11/2015).

3. Movimento 3 - Ocupar o Nicho

Passo 1a: Delinear os objetivos ou

Passo 1b: Apresentar a pesquisa ou

Passo 2: Apresentar as principais resultados ou

Passo 3: Indicar a estrutura do artigo.

(SWALES, 1990, p. 141.)

Mais recentemente, Swales e Feak (2009) apresentaram uma revisão simplificada do padrão anterior para:

1. *Background* / Introdução;
2. Propósito;
3. Metodologia / Tema / Procedimentos;
4. Resultados / Achados e
5. Discussão / Conclusões / Implicações / Recomendações.

Os modelos para as estruturas dos gêneros não são prescritivos, mas classificações para fins didáticos e, portanto, sujeitas a variações que advêm inclusive de características das diferentes áreas de pesquisa, como mencionado anteriormente. É importante ressaltar, de mais a mais, que inversões da ordem dos movimentos retóricos descritos são bastante frequentes nos *abstracts* dos diferentes campos de conhecimento. Já nos textos do *corpus* desta pesquisa, a variação de ordem é encontrada com frequência no acervo nacional. Além do que, introdução e propósito, muitas vezes, são o mesmo item.

Endres-Niggemeyer (1985) argumenta que os autores tendem a não seguir as indicações dos periódicos para redações de seus *abstracts* ou mesmo o padrão IMRD. Ela sustenta que os autores tendem a seguir seus próprios objetivos de formulação, uma vez que as orientações de publicação das revistas especializadas, muitas vezes, não vão ao encontro das necessidades dos leitores. Gledhill (2000) defende que as conclusões de Endres-Niggemeyer (*op. cit.*) são interessantes, porque sugerem que a estrutura retórica dos textos é menos marcada do que analistas de gênero, como Swales (1990), defendem. Ou seja, os

cientistas então incluíam seus próprios objetivos retóricos nos *abstracts*, em um estilo redacional mais livre do que o tradicionalmente esperado.¹⁰¹

Os textos acadêmicos, de acordo com Biber e Conrad (2009), não detêm características universais, podem variar situacionalmente (p. 3), inclusive tendo em vista suas condições de publicação. Zanella e Heberle (2015) acrescentam que “o contexto de situação pode influenciar a linguagem em que o *abstract* é escrito” (p. 17). Contudo, são os traços que podemos reconhecer como mais constantes que nos apontam o que seja mais relevante para a comunidade de usuários em questão. Do mesmo modo, tais traços são indicativos do que deve ser ensinado prioritariamente, como é interesse demonstrar em relação aos textos do acervo reunido para realizar o estudo.

Ao esforço de reconhecer os traços mais frequentes dos *abstracts*, Dayrell (2010) lembra que deve-se somar a necessidade de estar atento às peculiaridades culturais, visto que as práticas, expectativas e valores dos textos acadêmicos podem variar consideravelmente de uma língua para outra (CORTES, 2008; DAVOODIFARD, 2008 e DAYRELL, 2010/2012). A exemplo disso, destaca-se a probabilidade de apresentar as ações no começo das sentenças em inglês, diferentemente do português¹⁰². A variedade no uso de palavras de transição também é bastante enfatizada na tradição de pesquisa anglo-saxônica, enquanto em outras línguas, tende-se a recorrer, mais frequentemente, a repetições.

3.2.2. Dos estudos mais recentes que se relacionam com esta pesquisa

Chan e Foo (2004) identificam uma tendência mais holística e multidisciplinar no interesse em *abstracts* pelos praticantes de EAP, no século XXI. Como exemplo da atual multidisciplinariedade, citam-se também alguns estudos que se relacionam ao foco específico de nossa pesquisa.

Chris Gledhill, um dos investigadores que tem se dedicado a entender o padrão colocacional¹⁰³ dos gêneros acadêmicos, volta-se para *abstracts* no campo biomédico (1995;

¹⁰¹ Este recorte de pesquisa se volta para o estudo do padrão colocacional dos textos sem preocupação com a constituição específica da estrutura retórica nesses termos sequenciais.

¹⁰² A exemplo de ‘*Considering these results, we propose a probable mechanism of vitellogenin uptake regulation...*’ (https://www.researchgate.net/figure/51109988_fig5_Figure-5-Considering-these-results-we-propose-a-probable-mechanism-of-vitellogenin Último acesso em 03/04/16.)

¹⁰³ Relacionam-se à colocação, a co-ocorrência de duas (ou mais) palavras numa frequência maior do que seria de se esperar caso a co-ocorrência fosse aleatória (TAGNIN, 2004); aspecto que discutiremos no próximo capítulo.

1999; 2000, 2005, 2012). Dentre essas investigações, destacamos duas:

Gledhill 2005 apresenta uma visão sistêmica de adoção de itens lexicais em textos do campo da Farmácia, a fim de demonstrar como cientistas selecionam e apresentam informações em *abstracts* de artigos acadêmicos. Para tanto, esse estudioso alemão conjuga LC e o conceito de gêneros acadêmicos, argumentando que a primeira está em posição privilegiada de dar conta das preocupações tanto de analistas do discurso, quanto de aprendizes de EAP. Gledhill (*op. cit.*) procedeu ao que denomina uma análise de *corpus* sensível aos objetivos e práticas profissionais do tipo de texto analisado. O autor destaca que, embora se tenha conhecimento de muitas pesquisas voltadas para *abstracts* de textos científicos, ainda são relativamente poucas as que se voltam para a conjugação de gênero textual e fraseologia¹⁰⁴. O argumento é de que a investigação dos grupos de itens lexicais frequentes nos discursos acadêmicos cumpre papel importante em termos de definir escolhas retóricas efetivas. Para esse pesquisador, os padrões colocacionais representam um modelo implícito para as escolhas fraseológicas mais comuns, no sentido de que desvios no padrão habitual impactam os leitores e determinam mudanças nos modelos de pesquisa.

Na mesma direção, Gledhill (2012) analisa um *corpus* de 150 artigos sobre câncer, novamente, em busca de estabelecer padrões léxico-gramaticais. Nesse estudo, o investigador defende o trabalho com *corpus* especializado representativo de uma área de conhecimento como o mais adequado para o ensino de EAP, bem como, com vistas ao trabalho etnográfico com análise de gêneros acadêmicos de um modo geral.

Hyland e Tsé (2005) têm o mérito de estar dentre os primeiros a investigar *abstracts* produzidos por pesquisadores cuja primeira língua não é o inglês. Esses autores analisaram produções de acadêmicos chineses a partir de campos de conhecimento variados, explorando o uso do pronome relativo ‘*that*’ em dissertações e teses, enquanto a maior parte dos linguistas aplicados que tratam do gênero o fazem em contexto de artigo acadêmico exclusivamente.¹⁰⁵ Os autores constatam que, enquanto as construções que adotam ‘*that*’ para introduzir orações relativas são muito frequentes na tradição acadêmica anglo-saxônica, a exemplo de ‘*It’s possible to conclude that...*’, os autores chineses que observaram apresentavam muito pouca adoção desse pronome em seus textos. Essa interferência da língua

¹⁰⁴ De acordo com Blais (1993), assim como para Bevilacqua (2004), a noção de fraseologia abrange compostos, colocações, expressões idiomáticas, locuções, expressões fixas, co-ocorrentes e outras expressões do gênero, sendo uma das denominações adotadas para descrever os grupos mais ou menos fixos de itens lexicais nos textos especializados.

¹⁰⁵ Neste estudo, explora-se o gênero a partir de produções variadas, a fim de promover sistematização didática mais ampla, conforme será abordado no capítulo de metodologia.

chinesa na redação de artigos em inglês aponta um desvio no padrão usual e indica um aspecto importante para o ensino da língua inglesa entre pesquisadores chineses.

Carmen Dayrell, pesquisadora brasileira, também vem conjugando esforços para lidar com o tema, especialmente, cotejando *corpora* de aprendizes brasileiros a *corpora* de publicações internacionais (2008, 2009^a, 2009b, 2010, 2011, 2015), como forma de suporte na produção de materiais instrucionais ou de ferramentas computacionais. Abaixo, relacionamos algumas dessas investigações, as quais, fortemente, inspiram este estudo:

Dayrell e Aluísio (2008) exploram os padrões lexicais e as escolhas colocacionais em um *corpus* de *abstracts* de produções de pós-graduandos brasileiros, com base em associações adotadas para a palavra ‘*work*’, no sentido de ‘*trabalho acadêmico*’. Esse estudo teve o objetivo de prover *feedback* imediato aos participantes, bem como de chamar a atenção para os padrões lexicais mais frequentes para autores das áreas de Física, Farmácia e Ciências da Computação. De acordo com as pesquisadoras, questões relativas a escolhas lexicais são o ponto nevrálgico constatado no discurso de pós-graduandos brasileiros ao redigirem textos acadêmicos em inglês. A investigação constatou uma preferência, no *corpus* brasileiro, pelo uso da palavra ‘*work*’ com o sentido de pesquisa, enquanto, no *corpus* de pesquisas já publicadas, a preferência recaía sobre a palavra ‘*paper*’ e sobre ‘*study*’ no *corpus* inglês de comprovação do estudo. As diferenças reforçam a asserção tecida pelas autoras de que escolhas lexicais representam um desafio para os autores brasileiros.

Dayrell (2009a) explora as diferenças nas frequências e estruturas sintáticas de pacotes lexicais, a exemplo de ‘*a function of the, a wide range of, aim of this study, one of the most, the aim of this, the structure of the, the effect of the*’ relativamente ao mesmo grupo de aprendizes mencionado acima. A pesquisa apontou um uso bem mais acentuado dessas “sequências recorrentes de palavras associadas” (Hyland 2008, p. 5)¹⁰⁶ pelos falantes com menos domínio da língua do que pelos autores de publicações internacionais¹⁰⁷. Além do que, o *corpus* brasileiro ainda aponta preferências de associação de palavras distintas daquelas que são tradição internacional nas áreas de estudo dos pós-graduandos em questão.

Por fim, Dayrell *et al.* (2015) tem por objetivo averiguar o número de segmentos de palavras e o padrão léxico-gramatical empregados para especificação do *propósito* e do *resultado* em *abstracts* de alunos brasileiros cotejados a publicações

¹⁰⁶ “...*frequently recurrent strings of uninterrupted word-forms.*” (Hyland 2008, p.5).

¹⁰⁷ O estudo piloto desta pesquisa apontou a mesma tendência ao analisar a adoção de pacotes lexicais em no *corpus* nacional do acervo.

internacionais. Nesse caso, não foram constatadas diferenças significativas entre os dois *corpora*, no que se refere à expressão dos propósitos. No entanto, acerca dos resultados, foi percebido que os autores brasileiros, muitas vezes não os incluem, enquanto as publicações internacionais atribuem grande relevância a apresentá-los de modo detalhado. Esse trabalho é significativo, porque apresenta aplicação direta em sala de aula e para o aperfeiçoamento de ferramentas de escrita acadêmica.

Ao conjugarem esforços para contemplar questões de frequência de adoção de padrões lexicais e contexto de uso, as implicações didáticas dessas investigações inspiraram a construção deste estudo, cujo objetivo é entender o padrão colocacional de uma coleção de *abstracts* no campo biomédico.

Parte-se também de um alinhamento a Gledhill (2005) e Dayrell (2008, 2009^a, 2009b, 2010, 2011, 2015) por, igualmente, entender-se aqui que a investigação dos grupos de itens lexicais frequentes nos discursos acadêmicos cumpre papel importante em termos de definir escolhas retóricas efetivas. Em função disso, a investigação busca reconhecer as sequências mais constantes de palavras associadas que norteiam as escolhas lexicais de nosso *corpus*. Com isso, pretende-se dar suporte aos pesquisadores brasileiros, para que os *abstracts* que produzam estejam inseridos no padrão de qualidade léxico-gramatical das publicações internacionais de seu campo de atuação. Por fim, almeja-se que o estudo venha agregar uma contribuição aos esforços descritos neste capítulo, para fortalecer o papel do gênero textual que conjuga o máximo de informações com o mínimo de palavras.

4. CAPÍTULO 4

“Present real examples only. This is perhaps the first lesson to be learnt from corpus study. Language cannot be invented; it can only be captured.”

John Sinclair (1997, p. 31)

A LINGUÍSTICA DE *CORPUS*, OS ESTUDOS EM *ENGLISH FOR ACADEMIC PURPOSES* (EAP) E OS PACOTES LEXICAIS (*LEXICAL BUNDLES*)

Esta parte de nossa revisão de literatura apresenta três focos com perspectivas muito próximas: a Linguística de *Corpus* (LC), forma de aproximação da linguagem eleita para a investigação; os estudos em *English for Academic Purposes* (EAP), e a unidade analítica da pesquisa, os pacotes lexicais, também conhecidos, no Brasil, pela expressão inglesa *lexical bundles* (BIBER *et al.*, 1999).

4.1. A Contracultura da Linguística de *Corpus*

Embora o termo *Linguística de Corpus*, definido por McEnery e Wilson (1996) como o estudo da linguagem baseado em exemplos da vida real, tenha aparecido pela primeira vez apenas na década 80 (LEECH, 1992, p. 105), investigações baseadas em *corpora* têm uma história ancestral. Sabe-se que coleções de dados existiam desde a Grécia Antiga. De acordo com McEnery e Wilson (2001), entretanto, foi no começo do século XX que esse modelo de pesquisa começaria a ser adotado mais amplamente, na forma de análises de dados linguísticos examinados de forma manual, o que seria realizado, mais tarde, com algum apoio pré-computacional simples e suportes estatísticos. No final da década de 50, contudo, as críticas a essas investigações estatísticas foram tão severas que a LC se tornaria marginalizada (MCENERY *et al.*, 2006).

Nos anos 60, Noam Chomsky havia quebrado um padrão dos estudos linguísticos, ao postular que todos os dados necessários para o linguista estavam em sua própria mente e eram acessíveis por meio da introspecção. A concepção chomskiana “inaugurou a era do comprometimento total e incondicional da Linguística, com os preceitos cartesianos de fazer ciência” (RAJAPOGALAN, 2007, p. 25).

Estava lançada a Linguística Gerativa a partir de uma perspectiva racionalista, baseada na lógica do conhecimento; com ela, instaurou-se a caça ao empirismo. De acordo com essa visão, “os seres humanos são dotados de uma capacidade inata para a linguagem e possuem um conhecimento sobre o sistema linguístico, chamado competência” (KUPSKE, 2012, p. 176). A partir de então, os dados concretos seriam convocados apenas para

“confirmar um conceito universal da língua, para servir de exemplificações de modelos ou hipóteses previamente estabelecidos” (RAJAGOPALAN, 1998, p. 22).

As críticas originárias dos gerativistas encontraram eco tanto na supremacia do modelo racionalista, que havia desbancado o empirismo na época, quanto no argumento de que não parecia possível fazerem-se análises de dados extensos a olho nu, devido à grande probabilidade de erros. O advento tecnológico, contudo, impulsionaria uma virada na história dessa linha de estudos, fazendo com que a LC passasse a se orientar pela coleta e análise de conjuntos de dados computadorizados. A tecnologia da computação injetaria novo ânimo à Linguística, “em que, desde o princípio, imperou a tendência a teorizar a linguagem a partir do zero” (RAJAGOPALAN, 2007, p. 28).¹⁰⁸ É nessa conjuntura que o ano de 1964 veria nascer uma espécie de contracultura da Linguística.

O contexto histórico da organização do primeiro *corpus*¹⁰⁹ linguístico eletrônico, o *Corpus Brown*¹¹⁰ (FRANCIS e KUCERA, 1982), foi de efervescência para os estudos da linguagem. Justamente quando o mérito absoluto parecia pertencer ao terreno da competência, a Linguística de *Corpus* ressurgiu acenando com possibilidades inéditas para os estudos de ordem experimental.

Foi a necessidade que estudiosos como Firth (1957), Halliday (1991) e Sinclair (2003) sentiram de se apoiar em linguagem autêntica, para esboçarem generalizações ou teorias, que motivou epistemologicamente a constituição desse campo de estudos. Na forma de um movimento de resistência aos postulados gerativistas, portanto, seria lançada uma das áreas de pesquisa da linguagem mais ativas dos últimos anos. Rajagopalan (2007) arrisca a dizer que a ascensão da LC seja proporcional ao declínio das ideias do Gerativismo, o qual, em seu apogeu, impediu que outras formas de abordar a língua, que não através da introspecção, vigorassem.

O campo abriu caminho para muitas frentes de investigação, dentre as quais os estudos dos gêneros acadêmicos e o ensino de línguas adicionais. Embora, ainda hoje, restem críticas de parte daqueles que se ressentem de que falta-lhe a formulação de uma teoria a respeito do que é língua, em seu sentido mais abstrato, para dar sustentação às análises

¹⁰⁸ Vale ressaltar que os campos da LC e do Processamento de Línguas Naturais (PLN), este último voltado para a compreensão da língua, assim como para a organização de técnicas computacionais adequadas para o tratamento da língua escrita e falada (VIEIRA, 2004), são áreas distintas.

¹⁰⁹ Na definição de Kennedy, “nas ciências da linguagem, um *corpus* é um corpo de texto escrito ou de fala transcrita que serve como base para a análise ou descrição linguística”(1998, p. 1). A contribuição dessa definição é registrar explicitamente o objetivo de um *corpus*.

¹¹⁰ O *Brown University Standard Corpus of Present-Day American English* foi um *corpus* sincrônico lançado na Universidade de Stanford, Estados Unidos, em 1964, com um milhão de palavras, compiladas em inglês americano escrito, a partir de fontes variadas.

linguísticas. Entretanto, para seus entusiastas, a exemplo de Rajagopalan (2007), a quem recorreremos novamente,

Ela proporciona um olhar radicalmente distinto sobre a linguagem. Em vez de encarar a língua como algo pronto, acabado e hermeticamente fechado contra influências externas, como ensina a tradição estruturalista, a Linguística de *Corpus* a contempla como algo em construção, algo que está sendo constantemente trabalhado, aperfeiçoado (e adequado às nossas necessidades comunicativas do dia-a-dia), e, portanto, sujeito a modificações e inovações constantes. (RAJAGOPALAN, 2007, p. 23)

Sobre esse novo olhar, além das contribuições às pesquisas em EAP, a LC vem comprovando vigor ao servir de base para estudos originais em áreas variadas. Essas investigações abrangem desde a questão das mudanças climáticas (DAYRELL e URRY, 2015), passando por contabilidade e finanças (ATHANASAKOU, *et al.*, 2014), análise do discurso jornalístico (BAKER *et al.*, 2013), relações internacionais (GERMOND, 2010) e questões de gênero e sexualidade (BAKER, 2004), apenas para citar algumas das contribuições dos pesquisadores da Universidade de Lancaster/UK, onde, por exemplo, a LC vem sendo praticada de modo fértil.

4.1.1. Das concepções centrais

Os estudos em LC partem de uma visão sociolinguística da comunicação (FINATTO *et al.*, p. 159, 2010) e apresentam “um interesse pelo discurso, entendido como a linguagem em ação e em relação a seus usuários”¹¹¹ (GLEDHILL, 2000, p.10). Assim sendo, seu “objeto de estudo”¹¹² não é um fenômeno mental, mas um fenômeno social, algo observável e acessível através de evidências” (SHEPHERD, 2009, p. 151).

Duas concepções centrais são pilares do campo: a abordagem empirista e a visão da linguagem como sistema probabilístico. O princípio do empirismo se baseia em o conhecimento se originar da experiência, nesse caso, de dados que se organizam na forma de um *corpus*. A visão da linguagem como sistema probabilístico tem base nos postulados de Halliday (1991, 1992). Esse último entendimento pressupõe que “embora muitos traços

¹¹¹ “An interest in discourse (*language in action, language in relation to its users*).” (GLEDHILL, 2000, p. 10)

¹¹² Aspas da autora.

linguísticos sejam possíveis teoricamente, eles não ocorrem com a mesma frequência” (BERBER SARDINHA, 2000).

Segundo Finatto *et al.* (2010),

A língua, sob a ótica da LC, passa a ser entendida como um sistema probabilístico de combinatórias, de modo que uma dada palavra se define pela presença e pelos tipos de vinculação com as demais palavras dessa língua ao longo de todo um *continuum* de uso. (FINATTO *et al.*, 2010, p. 159)

De acordo com a base epistemológica do campo, os traços linguísticos não acontecem de modo aleatório e é possível apontar e quantificar padrões de regularidade, destacando uma correlação entre tais traços e os contextos situacionais de uso. A partir desses padrões, pode-se reconhecer que uma língua não se resume ao preenchimento de espaços vazios de forma arbitrária. O ambiente linguístico, denominado de cotexto, atua sobre a co-seleção de itens lexicais, ou seja, um determinado item apresenta preferência por outro, em determinado ambiente linguístico.

É por isso que Finatto *et al.* (2010) identifica, dentre os princípios básicos da LC, “uma configuração sistêmico-funcional do significado no contexto” (p. 159). Isso significa que o princípio é *sistêmico* porque a língua é vista como um sistema de escolhas potenciais não-arbitrariamente motivadas e *funcional* porque procura explicar as implicações comunicativas de uma seleção dentro desse sistema de escolhas.

Esses aspectos remetem à questão dos padrões de uso e, portanto, aos princípios idiomáticos que foram postulados por Sinclair (1991). De acordo com essa visão, a linguagem funciona em blocos fixos e/ou relativamente fixos e se orienta por princípios maiores do que a livre escolha dos usuários. Tomando por base um exemplo do *corpus* desta investigação, a saber, ‘*the end of*’ é uma sequência cuja continuidade é limitada por uma próxima palavra da categoria de um artigo. Nos dados do estudo, os artigos que se seguem à ‘*the end of*’ são, majoritariamente da classe dos definidos, confirmando uma *preferência* dos gêneros/registros textuais acadêmicos (HYLAND, 2008; BIBER *et al.*, 1999) por uma maior incidência dessa associação de palavras. Assim, a sequência ‘*the end of*’ se define por uma forte associação com a partícula ‘*the*’.

Segundo o princípio da livre escolha, por outro lado, seria possível selecionar, apenas com restrições gramaticais, os itens que compõem um enunciado. Em termos de discursos acadêmicos, isso implicaria o estabelecimento de associações lexicais pelos autores dos textos independentemente do que o campo de conhecimento, o ambiente linguístico, já

tenha como estabelecido para a área. Se aceitássemos essa possibilidade, descaracterizaríamos as associações lexicais pelas quais é possível identificar um campo de conhecimento. Dito de outro modo, desprezaríamos a concepção de que o caráter recorrente de um registro pode ser identificado como traço comum a uma variedade acadêmica, porque re-ocorre nas escolhas léxico-gramaticais dos usuários fluentes (BIBER, *et al.*, 1999) ¹¹³.

Esse postulado estabelece que as escolhas linguísticas envolvem a seleção simultânea de estruturas maiores, nas quais as palavras não estariam, no caso, sujeitas à discriminação do escritor fluente no domínio de um gênero textual. A recorrência acontece na forma de cadeias lexicais, também denominadas de *n-gramas*¹¹⁴ ou *clusters*, que podem ser formadas por 2, 3 ou mesmo mais de 5 elementos¹¹⁵. Ao residir no princípio idiomático, a proposta de Sinclair possibilita a expansão dos padrões fraseológicos de uso da linguagem.

A partir do princípio idiomático, ressaltam-se as seguintes formas de investigações que abordam a tendência de certas palavras para se associarem umas às outras:

- 1- Colocações¹¹⁶ – associações entre itens lexicais ou entre o léxico e os campos semânticos, o que remete a um padrão de uso voltado para o léxico cotejado às escolhas feitas pelos usuários de uma língua, a exemplo de “*results of + sensitivity analysis*” ou “*it is + possible that*”;
- 2- Coligações – associações de itens lexicais em posições estruturais que se baseiam em ligações de co-ocorrência positiva e remetem à ocorrência de uma classe gramatical ou padrão estrutural com outro, ou mesmo com uma palavra ou frase (SINCLAIR, 2003, p. 173), a exemplo de ‘*as*’ ser mais usado com presente perfeito em inglês acadêmico, a exemplo de ‘*as shown /*

¹¹³ Lembrando uma consideração feita na pág. 12 do cap. 2 a esse respeito, Hoffmann (2015) reforça, desde longa data, que os textos especializados apresentam, em média, sessenta por cento de seu corpo composto por substantivos e adjetivos. Esses se sobrepõem às outras classes de palavras, com vistas a designar os objetos da atividade especializada (p. 43-4). Na verdade, acrescenta-se que Kennedy (1998) também observa que a frequência de substantivos na língua inglesa corresponde a cerca de 25%, sendo maior do que as demais categorias. Biber *et al.* (1999) já havia registrado a ocorrência de 300 mil substantivos por milhão de palavras na prosa acadêmica escrita. Posteriormente, Biber e Gray (2013^b) apresentam entendimento semelhante ao de Hoffmann (OP.CIT.) em estudo acerca da nominalização da escrita.

¹¹⁴ *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).

¹¹⁵ Esta investigação explora esses padrões na forma de associações de palavras denominadas por Biber de *pacotes lexicais* (1990, p. 262), que apresentaremos na seção 4.3.

¹¹⁶ Este padrão foi originalmente proposto por Firth (1957).

seen / demonstrated’;

- 3- Preferências Semânticas – preferências de certos itens lexicais por áreas conceituais, a exemplo de ‘ago’ que se relaciona a expressões temporais, como em ‘*three year ago*’, ‘*a long time ago*’, ‘*a week ago*’ ou o substantivo ‘*meal*’ que se relaciona com os verbos ‘*have*’ ou ‘*eat*’;
- 4- Prosódias Semânticas¹¹⁷ – preferências de certos itens por prosódias negativas ou positivas, a exemplo de ‘*cause*’, que apresenta conotação negativa: ‘*cause damage*’, ‘*cause pain*’, ‘*cause distress*’, ‘*cause concern*’ (XIAO e MCENERY, 2006, p. 114)¹¹⁸;
- 5- Associações entre Itens Lexicais e seus Contextos – tendência dos itens lexicais a acontecerem em determinados contextos com frequência específicas (SHEPERD, 2009, p. 155), a exemplo de ‘*as shown in the figure*’ e ‘*with respect to the*’, que tendem a aparecer em gêneros textuais acadêmicos (BIBER *et al.*, 1999, p. 997-1025).

É com base nesses padrões que Berber Sardinha (2000) afirma que, “quando se diz que a variação não é aleatória, na verdade, está-se afirmando que a linguagem é padronizada” (p. 351), padronização esta que se dá pela recorrência, isto é, “uma colocação, coligação ou estrutura, que se repete significativamente, mostra sinais de ser, na verdade, um padrão lexical ou léxico-gramatical.

Uma das grandes contribuições da LC para os estudos linguísticos foi a constatação empírica de que se pode encontrar, através de busca eletrônica, *n*-gramas ou agrupamentos de itens lexicais, em qualquer conjunto de dados submetidos a programas eletrônicos. São esses conjuntos de dados que nos fornecem evidências a respeito do princípio idiomático adotado pelos usuários. Para que seja possível verificar quais são os agrupamentos lexicais mais frequentemente empregados pelos falantes ou atestados pelo uso, o trabalho com

¹¹⁷ No caso de Preferência Semântica, está-se tratando do significado dos colocados, sendo a relação entre palavras o aspecto mais importante. No caso de Prosódia Semântica, quer-se abordar o significado efetivo de uma palavra em relação a seus colocados. Embora os conceitos sejam diferentes, eles são interdependentes (STUBBS, 2001).

¹¹⁸ Embora também seja usual dizer-se ‘*cause happiness*’.

base em *corpus* mostra-se fundamental.

4.1.2. Do trabalho de *corpus*

Dentre as vantagens de adoção de *corpora* para a pesquisa linguística estão as possibilidades de explicar as diferenças nos usos das palavras e de formas linguísticas, dentre outros traços, com base na probabilidade de ocorrência em determinados contextos (BIBER *et al.*, 1998). Esses fatos não poderiam ser explicados apenas pela intuição (SINCLAIR, 1991). Até porque, conforme salienta Sheperd (2009), a empreitada analítica “não pode depender das intuições do pesquisador, visto que os seres humanos tendem a reconhecer aquilo que não é típico com mais frequência do que aquilo que é padronizado” (p. 152). Os *corpora*, portanto, são utilizados para gerar conhecimento empírico sobre as línguas.

De acordo com Sinclair (1991), “um *corpus* corresponde a uma coletânea de textos naturais escolhidos para caracterizar um estado ou variedade de linguagem” (p. 171). Já nos termos de Biber *et al.* (1999), a noção de *corpus* é abordada, naturalmente, pela perspectiva de registro: “uma coleção de textos falados ou escritos, organizada por registro e codificada para outras considerações discursivas, abrange um *corpus*”¹¹⁹ (p.4). A literatura, entretanto, registra muitas definições para o que seja *corpus* (a exemplo de ATKINS, CLEAR e OSTLER, 1992; FRANCIS, 1992; KENNEDY, 1998 e MCENERY *et al.*, 2006) O consenso é que deva ser composto de:

1. Dados Linguísticos Autênticos;
2. Segmentos Legíveis por computador;
3. Porções de Língua Especialmente Organizadas;
4. Textos Capazes de Representar uma Língua ou Variedade de Língua Particular;

(MCENERY *et al.*, 2006, p. 5.)

A respeito dessa descrição, quando mencionam-se (1.) dados linguísticos

¹¹⁹ “A collection of spoken and written texts, organized by register and coded for other discourse considerations, comprises a corpus” (BIBER *et al.*, 1999, p. 4).

autênticos, quer-se dizer que esses não devem ser inventados. Do contrário, os dados não desfrutariam de legitimidade para estudos acerca de frequência ou particularidades de uso de linguagem. Isso posto, é importante discutir o que é entendido aqui como texto.

Berber Sardinha (2000) relaciona que a delimitação do conceito de texto pode ser problemática em relação ao trabalho de *corpus*, já que podemos estar considerando “tanto um artigo científico, quanto o seu resumo inicial, quanto um trecho de conversação, como texto”(p. 336)¹²⁰. Por isso, esse autor refere que o uso do termo *porções de linguagem* pode ser mais adequado. Além da questão da extensão dos itens da coleção de dados, há outro aspecto já mencionado anteriormente também preponderante. O que quer que se considere texto, *vis-à-vis* um *corpus*, é submetido a uma transformação em função da análise linguística efetuada pelas ferramentas computacionais. Finatto (2011^a, p. 8) pondera que, para as análises de LC, o texto transforma-se em *corpus*. Nessa instância, o mesmo pode perder as características do todo para se submeter a microanálises de suas peculiaridades. Desse modo, reforça-se o entendimento de Berber Sardinha (*op. cit.*), quando esse expoente da LC brasileira destaca que a noção de texto em relação à *corpus* é problemática ou, minimamente, deve ser observada com cautela.

Em relação à (2.) legibilidade dos dados por meio de computador destacada por McEnery *et al.* (2006) acima, pesa aqui o fato de essas análises serem, não apenas ágeis, mas precisas. Tais aspectos contrastam com o caráter dos estudos em LC da primeira metade do século XX. Entram em cena os programas adotados para análise de *corpus* que podem ser ferramentas já existentes ou mesmo desenvolvidas especialmente para os propósitos de uma investigação. Nesse sentido, as contribuições da Linguística de base Computacional, no desenho de novas ferramentas, têm sido grandes aliadas da LC, a exemplo dos trabalhos de Anthony *et al.* (2011), Hardie, (2005; 2007); Rayson *et al.* (2004) e Leech *et al.* (1994), dentre os de âmbito internacional.

O fato de os dados compilados para integrar um *corpus* serem apresentados como (3.) *organizados* para esse fim implica que representam um recorte ou que são destacados para uma finalidade analítica, mais especificamente, para a pesquisa linguística. Nas palavras de Berber Sardinha (2000), “o *corpus* em si é artificial, um objeto criado com fins específicos de pesquisa” (p. 336).

Por fim, por (4.) *representação específica de uma língua ou uma variedade de*

¹²⁰ Habert *et al.* (2001), em um artigo acerca da construção de um *corpus* representativo da linguagem biomédica, defende, por exemplo, a adoção de textos compactos, a exemplo de *abstracts* para formação de *corpus* na área (p. 245).

linguagem quer-se dizer que é necessário que o corpus contenha traços comparáveis entre si, tais como textos acadêmicos ou textos históricos, por exemplo. Por outro lado, não há consenso acerca que seja de fato *representativo*.

Para Sinclair (1996), um *corpus* deve ser o maior possível “dentro do que pode ser atingido com a tecnologia da época” (p. 3), para que seja representativo. Já McEnery (2015) pondera que “o tamanho de um *corpus* está longe de ser uma questão tranquila para as pesquisas em LC. São as perguntas de pesquisa que vão determinar se a extensão do acervo é adequada para dar conta de analisar um fenômeno linguístico”¹²¹. De todo modo, Aston (1997) considera que um *corpus* pequeno abrange de 20 a 200 mil palavras, enquanto um *corpus* grande seria composto de 100 milhões ou mais.

Finatto *et al.* (2010), ao ponderar que “a LC é marcada pela observação *o mais extensiva*”¹²² possível, dos usos reais da língua” (p.158) estabelece que

o importante é reconhecer, nos *corpora*, a variedade extensiva dos desempenhos linguísticos com vistas à depreensão *a posteriori* de padrões. As fontes de dados devem permitir a identificação de padrões e de especificidades dos usos da língua em diferentes situações, inclusive na macro-situação dos textos técnico e científicos. (FINATTO *et al.*, 2010, p. 158)

Outro aspecto atribuído à representatividade de um *corpus* é a *especificidade*. Para que um *corpus* seja representativo, é necessário que ele seja específico (MCENERY e WILSON, 2001). Ao se adotar uma variedade específica da linguagem para organizar uma coleção de dados, a exemplo deste trabalho que congrega *abstracts* do campo temático da saúde, atinge-se representatividade através da especificidade de um léxico que comporta menos variação e, conseqüentemente, uma maior padronização.

Por fim, é fundamental que o conjunto de dados organizado para uma pesquisa seja adequado para o desenvolvimento da investigação. Hasan (1992) se refere ao caráter adequação em termos de o *corpus* “estar afinado com os objetivos da análise” (p. 301). O *corpus*, portanto, precisa dar conta das perguntas de pesquisa, o que faz com que Berber Sardinha (2000) questione o caráter adequação dos ‘*corpora prêt-à-porter*’, já que esses são criados para representar a língua como um todo (p. 349). No caso deste estudo, o *corpus* foi compilado especialmente para dar conta dos objetivos da investigação, para que fosse dado conhecer as características de textos acadêmicos da área da saúde e, com isso, lidar com os questionamentos estabelecidos acerca do funcionamento da linguagem nesse segmento

¹²¹ Comunicação pessoal em Lancaster University, 01/06/2015.

¹²² Grifo meu.

específico de realização.

Ao longo do advento dos grandes *corpora* eletrônicos, três deles representam um marco histórico, sendo todos esses de língua inglesa: o *Corpus Brown*, em função do pioneirismo; o *Corpus Birmingham*¹²³, por ter sido o primeiro a ultrapassar a marca de um milhão de palavras e o *British National Corpus* ou *Corpus BNC*¹²⁴, por ter sido o primeiro a atingir a dimensão de cem milhões de palavras. Não obstante, destacam-se o *Corpus LOB*¹²⁵ e o *Corpus London-Lund*¹²⁶, pois, com base nesses *corpora*, Douglas Biber (1988) desenvolveu sua meticulosa descrição dos gêneros da língua inglesa a que referimos anteriormente.

Alguns *corpora* importantes que se encontram disponíveis *online* são os seguintes¹²⁷:

Tabela (1): Corpora em Língua Inglesa Disponíveis Online

<i>Corpus</i>	Número de Palavras	Língua/ Variante	Período
<i>Wikipedia</i> http://corpus.byu.edu/wiki/	1.9 bilhões	Inglês	2014
<i>Global Web-Based English</i> (<i>GloWBE</i>) http://corpus.byu.edu/glowbe/	1.9 bilhões	20 países	2012-2013
<i>Corpus of Contemporary American English</i>	450 milhões	Inglês Americano	1990-2012

¹²³ O *Corpus Birmingham* foi desenvolvido na década de 80, contendo 17 milhões de palavras, originárias do inglês escrito.

¹²⁴ O *British National Corpus (BNC)* é uma coleção de dados que congrega linguagem oral e escrita, originária de um amplo número de fontes. Foi concebido para representar uma extensa variedade de inglês britânico no final do século XX. Disponível em <http://www.natcorp.ox.ac.uk> (Último acesso em 27/11/2015)

¹²⁵ O *Lancaster-Oslo/Bergen Corpus (Corpus LOB)* é um *corpus* de 1 milhão de palavras de inglês britânico, que foi compilado em 1970, em colaboração entre a Universidade de Lancaster, a Universidade de Oslo e o Centro de Computação para Humanidades Norueguês, para apresentar uma contrapartida britânica ao *Corpus Brown*.

¹²⁶ O *London-Lund* é um *corpus* de inglês oral organizado em 1990, pela Universidade de Londres em conjunto com a Universidade de Lund na Suécia.

¹²⁷ Para uma relação mais extensiva, abrangendo *corpora* com concordanciadores e algumas ferramentas disponíveis online vide <http://courses.washington.edu/englhtml/engl560/corplingresources.htm> (Último acesso em 27/11/2015).

(COCA) <i>http://corpus.byu.edu/coca/</i>			
<i>Corpus of Historical American English</i> (COHA) <i>http://corpus.byu.edu/coha/</i>	450 milhões	Inglês Americano	1810-2009
<i>TIME Magazine Corpus</i> <i>http://corpus.byu.edu/time/</i>	100 milhões	Inglês Americano	1923-2006
<i>Corpus of American Soap Operas</i> <i>http://corpus.byu.edu/soap/</i>	100 milhões	Inglês Americano	2001-2012
<i>British National Corpus</i> (BYU-BNC) <i>http://corpus.byu.edu/bnc/</i>	100 milhões	Inglês Britânico	1980-1993
<i>Straty Corpus</i> (Canadá) <i>http://corpus.byu.edu/can/</i>	50 milhões	Inglês Canadense	1970-2000

Fonte: Freitas (2016)

A partir da pesquisa com *corpora* organizados em computadores, a implementação de muitos trabalhos se tornou realidade. Como exemplo de materialização desses tem-se dicionários como o *Collins Cobuild English Dictionary* (1987), a *Longman Grammar of Spoken and Written English* (BIBER *et al.*, 1999) e mesmo de ferramentas de redes lexicais, como a *WordNet*¹²⁸.

¹²⁸ <https://wordnet.princeton.edu> (Último acesso em 25/11/2015.)

Existe uma grande variedade de tipos de *corpora* na classificação dos especialistas (JONES e WALLER, 2015; TAGNIN, 2010; BERBER SARDINHA, 2000).

Um *corpus* pode variar especialmente em relação a:

1. Tipologia – falados ou escritos;
2. Temporalidade – sincrônicos (a partir de um período de tempo), diacrônicos (cotejando períodos diferentes), contemporâneos (da atualidade) ou históricos (de um período findo);
3. Seleção - de amostragem (a partir de variedades textuais), monitor (o conteúdo foi atualizado), dinâmico (novos textos vão sendo incluídos continuamente), estático (a coleção de textos não sofre acréscimos) ou equilibrado (os textos são distribuídos de forma proporcional);
4. Conteúdo – geral ou especializado (de gêneros ou registros definidos), regional (de variedade(e) sociolinguística(s) específicas) ou multilíngue (inclui línguas diferentes);
5. Autoria – de aprendiz (os autores estão desenvolvendo o domínio da língua), de língua nativa (os autores não estão aprendendo a língua de comunicação);
6. Disposição Interna – paralelo (originais e respectivas traduções) ou alinhado (as traduções aparecem abaixo do original);
7. Finalidade – de estudo (que se pretende descrever), de referência (balizado em relação ao de estudo) ou de treinamento (criado para aplicação de ferramenta de análise);
8. Informação Linguística – anotado¹²⁹ (adicionam-se informações codificadas ao texto) ou não-anotado.

¹²⁹ De acordo com Teixeira (2007), anotação ou etiquetagem é a inclusão automática, semi-automática ou manual de qualquer tipo de informação em um *corpus* de estudo, com vistas a facilitar a análise. A autora apregoa que a etiquetagem pode ser de vários tipos: morfossintática (que indica a classe gramatical de cada palavra do *corpus*), sintática (que analisa as relações formais entre os elementos das sentenças) e semântica (que caracteriza as palavras com base no sentido). (p. 117-118)

Para a realização deste trabalho, por exemplo, adotou-se um *corpus* escrito contemporâneo (já que os textos acadêmicos que o integram são atuais), composto de uma coletânea de *abstracts* originários de 4 universidades brasileiras e de três revistas especializadas internacionais. O conjunto de dados foi organizado de modo equilibrado, uma vez que tem-se um número de palavras quase idêntico no *subcorpus* brasileiro e no internacional. Findo este estudo, é intenção que o acervo passe a ser dinâmico, sofrendo acréscimos, para que seja possível tecer novas análises que possam enriquecer a criação de futuros materiais de ensino. Os dados são de natureza especializada, pois adotou-se um gênero/registo definido, além disso, são monolíngues, pois apresentam-se apenas em língua inglesa. Quanto à questão de autoria, para fins da descrição acima, usou-se o parâmetro de um grupo de autores consagrados no campo. Todavia, é mais politicamente contemporâneo para as concepções da Linguística Aplicada descrever o *corpus* como parte de usuários de inglês como língua franca (critério adotado para nomear a coleção de *abstracts* brasileiros) e, ainda, parte de usuários experientes (critério adotado para designar a parte da coleção que se origina dos periódicos internacionais). Isso porque não se conhecem as origens dos autores dos textos das publicações que figuram no *subcorpus* internacional, mas o padrão de qualidade estabelecido para publicações especializadas de alto impacto determina que esses textos apresentem um caráter léxico-gramatical de excelência. Por fim, os dados organizados integram um *corpus* de estudo, uma vez que ambos os conjuntos de dados serão descritos, embora o *corpus* internacional seja de referência em relação ao nacional. Para fins deste trabalho, em função do recorte determinado pelas perguntas de pesquisa, optou-se por não incluir anotação. Contudo, o trabalho de adicionar informações morfo-sintáticas, sintáticas ou semânticas em forma de códigos-padrão inseridos em documento eletrônico, é capaz de agregar valor aos dados, uma vez que permite lidar com uma amplitude de indagações analíticas.

4.1.3. Das análises computadorizadas e do principal objetivo das investigações: ‘diga-me com quem andas’

Na década de 60, apenas as universidades contavam com computadores. As máquinas eram do tipo *mainframe*, e a pesquisa de *corpus* ressurgiu restrita a esses centros. Foi com a chegada dos computadores pessoais, a partir da década de 1980, que os *corpora*

realmente se popularizaram. Dada a capacidade de armazenamento e processamento de dados, *corpora* maiores, assim como mais ferramentas de pesquisa, foram sendo desenvolvidos. Passou-se a observar fenômenos de linguagem que não seriam percebidos se o número de dados fosse pequeno. Do mesmo modo, passou-se a “poder quantificar a linguagem ou, dito de outra maneira, calcular frequências que indicam não o que é possível, mas sim o que é mais provável na linguagem” (JACOBI, 2001, p. 3). Além do mais,

Com a popularização dos computadores, foi possibilitado o acesso de mais pesquisadores ao processamento de linguagem natural e, concomitantemente, a sofisticação do equipamento permitiu a consecução de tarefas mais complexas, mais eficientemente, sem falar no aumento da capacidade de armazenamento e na introdução de novas mídias (fitas magnéticas, em vez de cartões *hollerith* perfurados, etc.), as quais facilitaram a criação e manutenção de *corpora* em maior número. (BERBER SARDINHA, 2000, p. 327)

Portanto, o fato de serem analisáveis por máquinas é um atributo dos *corpora* modernos, em função da velocidade de processamento e da precisão. Além disso, em alguns casos, o trabalho não envolve custos com a ferramenta de processamento¹³⁰.

As ferramentas eletrônicas são utilizadas para reorganização e extração de informações dos *corpora* para fins de análise linguística, podendo utilizar técnicas automáticas e interativas. Dentre as ferramentas computacionais mais comuns para esses estudos estão:

- a. Programas para Listar Palavras – executam a contagem de palavras de um *corpus*;
- b. Programas Concordanciadores – permitem a busca por palavras específicas em um *corpus*, fornecendo listas para as ocorrências em contexto na forma de uma concordância;
- c. Programas Etiquetadores – fazem análises automáticas de um *corpus* e inserem etiquetas (códigos) de ordem morfossintática, sintática, semântica ou discursiva.

Um *corpus*, portanto, é normalmente mensurado pelas ferramentas eletrônicas de análise em relação ao seu número de palavras. Na terminologia da LC, diz-se que o *corpus* é mensurado em relação ao número de *tokens* (o número total de itens ou palavras, incluindo as repetições de um mesmo item ou palavra) ou mesmo em relação ao número de *types* (cada

¹³⁰ Este é o caso da ferramenta adotada para desenvolvimento deste trabalho, o *AntConc* (ANTHONY, 2014), que será apresentada no capítulo de metodologia e que é disponibilizada pelo autor de forma gratuita, o que democratiza e promove o trabalho com *corpora*.

item ou palavra, sem considerar as repetições). Os *types* podem ser contabilizados de forma lematizada (a partir da forma básica de uma palavra) ou não. Pode-se também relacionar a *type/token ratio* ou relação entre *type* e *token*, que aponta a riqueza lexical do texto (quanto maior a relação, mais variado é o léxico do texto).

O trabalho com frequência se tornou viável, especialmente, em função das análises estatísticas, que podem fornecer indícios acerca de muitos aspectos da língua. A descoberta de que a linguagem está padronizada tanto lexical quanto gramaticalmente baseia-se em análise estatística de concordâncias, que são listas de orações, completas ou não, extraídas de um *corpus*, nas quais uma ou mais palavras aparecem centralizadas. A palavra central, que é significativamente mais frequente, quando comparada com as demais, recebe o nome de nóculo ou palavra-chave (também denominada de *key word* ou *cluster*). O texto à esquerda e à direita do nóculo é chamado de co-texto, como no exemplo a seguir a partir do *AntConc*, em que o nóculo é a palavra ‘*research*’, em um exemplo a partir de nosso *corpus*:

Figura (1): Linhas de Concordância – AntConc Versão 3.2.1



Fonte: Freitas (2016)

As concordâncias possibilitam a análise de um grande número de dados a partir dos quais é possível calcular as frequências de co-ocorrência de palavras. A visibilidade é uma característica importante das concordâncias e é consequência da disposição das

ocorrências com a palavra-chave centralizada. Esse tipo de disposição e a possibilidade de reordenar alfabeticamente a listagem facilitam a observação das palavras que ocorrem antes e depois do nóculo. Desta forma, é possível revelar padrões que se repetem e que não seriam observáveis com um número pequeno de dados ou a olho nu. Tais concepções, somadas aos recursos analíticos, tiveram consequências diretas, especialmente, para o ensino de línguas.¹³¹

Com base nas asserções feitas, chega-se ao principal objetivo das investigações que adotam a LC, que é estudar os padrões e as variações sistemáticas que a língua apresenta.

Conforme Berber Sardinha (2004), “de um modo geral, a padronização é a regularidade expressa na recorrência sistemática de unidades co-ocorrentes de várias ordens (lexical, gramatical, sintática, etc.)” (p. 40). Para que sejam definidos os padrões de uma palavra, faz-se necessário averiguar as outras palavras e estruturas frequentemente associadas a ela que, de alguma forma, refletem o seu significado. Isso representa uma espécie de ‘diga-me com quem andas e te direi quem és’ léxico-gramatical. É por isso que Sinclair postulou que *se conhece uma palavra pelas companhias com as quais ela anda* (FIRTH, 1957, p. 11).

Como as pesquisas em LC estão interessadas em padrões linguísticos a partir de dados empíricos e no que esses revelam sobre o uso, apresentam uma visão mais objetiva sobre o comportamento da linguagem do que as investigações que adotam a introspecção e a intuição. Uma análise que adota *corpus* está em posição de investigar praticamente qualquer padrão de linguagem, quer seja estrutural, léxico-gramatical, discursivo, fonológico ou morfológico. Mesmo a adoção de agendas muito específicas como questões sociolinguísticas acerca de uso de língua por homens ou mulheres (MCENERY *et al.*, 2004) ou a forma como a imprensa americana lidou com as vítimas do Furacão Katrina (POTTS *et al.*, 2015) podem ser objeto de investigação por essa linha de estudos. Com a adoção de ferramentas analíticas adequadas, é possível ao investigador revelar padrões de uso da linguagem. Também é possível verificar a extensão com que os padrões são adotados e os fatores contextuais que influenciam na variabilidade.

Por outro lado, dizermos que a LC está em posição de revelar padrões de frequência e variações sistemáticas não implica estar-se em posição de afirmar que algo esteja *correto* em termos de padrão léxico-gramatical. Dessa forma, os *corpora* são capazes de nos fornecer evidências, mas não explicações ou informações acerca do funcionamento da linguagem.

¹³¹ Flowerdew e Forest (2009), por exemplo, em estudo a respeito da realização das estruturas esquemáticas e léxico-gramaticais do discurso acadêmico, verificam que, enquanto em ciências humanas, as palavras-chave tendem a ser abstratas, nas ciências naturais, as mesmas tendem a ser mais marcadamente conceituais, revelando, portanto, um discurso de natureza mais conceitual.

Por outro viés, pode-se também depreender que o trabalho de *corpus* reserva lugar de destaque à interpretação do investigador. Com isso, nós, pesquisadores, temos parte fundamental em interpretar as evidências empíricas. Por extensão, não obstante o número de ocorrências de um padrão de uso da linguagem em um *corpus*, sempre haverá subjetividade na forma como analisamos ou compreendemos os dados ou no modo como estabelecemos cruzamentos com outras perspectivas investigativas. Dito de outra forma, as pesquisas em LC não se organizam apenas pelo teor quantitativo, pois o posicionamento do analista com base, inclusive, em suas concepções acerca do tema em estudo, são fundamentais para as conclusões a respeito dos modos de uso da linguagem. Esse aspecto, relacionado às pesquisas em LC, torna sua adoção particularmente frutífera.

Outrossim, em função de o trabalho de *corpora* transformar os textos no momento da análise, o cotejamento da linguagem em relação a seu contexto original (a exemplo de em relação a imagens ou figuras) não é possível. Igualmente, “nenhum *corpus* consegue incorporar a totalidade de uma língua, independentemente do número de palavras que contenha”¹³² (JONES e WALLER, 2015, p. 15).

4.1.4. Dos tipos de estudos em LC

Tagnin (2010) explica que existem dois tipos de estudos em LC: os do tipo que se baseiam em *corpus* (*corpus-based*) e aqueles que são direcionados pelo *corpus* (*corpus-driven*). Os do primeiro estilo adotam o *corpus* para (a) comprovar (ou não) uma hipótese ou (b) para extrair exemplos, enquanto aqueles que se estabelecem direcionados pelo *corpus* se orientam pelos dados, sem pressuposições teóricas (p. 357-358).

Correntemente, muitos investigadores do campo de EAP no contexto de ensino de línguas adicionais desenvolvem estudos baseados em *corpora* (dentre esses, CONNOR *et al.*, 2002; BHATIA *et al.*, 2004; FLOWERDEW, L. 2008), para confirmar crenças ou para desconstruir suposições a respeito de padrões de uso da linguagem.

Já os trabalhos direcionados por *corpora* tiveram o mérito de impulsionar a criação de uma importante teoria, denominada Teoria do *Priming Lexical*, de acordo com a qual os papéis desempenhados pelo léxico e pela gramática se revertem (HOEY, 2005). O

¹³² “No corpus can incorporate the totality of language, however many words it contains.” (JONES e WALLER, 2015, p. 15)

princípio é de que o léxico é orientado de forma organizada e complexa, e a gramática é o resultado da estrutura lexical (SHEPHERD, 2009, p. 155).

Para esta pesquisa, adota-se a segunda categoria de estudo, já que, para desenvolvê-la, não se partiu de hipótese prévia acerca das principais associações lexicais a serem verificadas na coleção de dados. Entretanto, naturalmente, há uma expectativa de encontrar elementos fixos e variáveis a serem identificados, embora não se saiba, *a priori*, quais são as unidades lexicais significativamente mais frequentes no *corpus* ou como essas se comportam estrutural e funcionalmente. Quer-se dizer que o pesquisador sempre parte de um determinado ponto relativo a sua visão do funcionamento da língua, a partir de um entendimento determinado. Não obstante, neste caso específico em que se quer conhecer a composição dos pacotes lexicais mais frequentes de um *corpus* no campo da saúde, é a partir de sua comprovação de frequência pelo uso da ferramenta eleita que saberemos exatamente o que será prioridade categorizar e ensinar.

O’Keeffe e Walsh (2012), por sua vez, vão além da tipologia de estudos baseados ou direcionados por *corpus* ao estabelecerem que, conjugadas a essas categorias, as pesquisas em LC podem ser de duas outras ordens ainda:

1. Descritivas – Investigações em que o *corpus* é um fim em si. Nesse caso, o pesquisador analisa os dados para escrutinar o emprego da linguagem e alargar as descrições dos padrões de uso em um gênero discursivo particular, a exemplo do estudo de Bednarek (2006), em que a pesquisadora avalia o padrão linguístico de um *corpus* de jornais. Ou seja, o objetivo é conhecer a linguagem que tipifica o acervo de dados;

2. Aplicadas – Investigações em que o *corpus* serve como meio para atingir um fim. Nesse modelo, o pesquisador mira além dos dados para tratar tanto das perguntas de pesquisa, quanto da análise. O *corpus* é uma ferramenta metodológica que conduz a análises em combinação com outras abordagens teóricas, a exemplo de O’Halloran (2010) que explora um *corpus* de artigos de jornais britânicos sobre o tema imigração. O estudo estabelece um cruzamento com a Análise Crítica do Discurso¹³³ e o resultado é uma profunda análise entre uso de linguagem e padrões ideológicos.

Já Kennedy (1998) percebe que a multiplicação dos trabalhos, tanto baseados

¹³³ A Análise Crítica do Discurso é uma abordagem transdisciplinar ao estudo dos textos, que considera a linguagem como uma forma de prática social. Para conhecer o campo, vide Fairclough (1989).

em *corpus*, quanto direcionados por *corpus*, determina um crescimento de quatro campos de pesquisa:

- (1) compilação de corpus
- (2) desenvolvimento de ferramentas
- (3) descrição da linguagem
- (4) aplicação de corpora (ensino de línguas, reconhecimento de voz, tradução, etc.)

(KENNEDY, *op. cit.*, *apud* BERBER SARDINHA, 2000, p. 358)

A respeito dos trabalhos que se valem de *corpus*, é fundamental acrescentar ainda a mudança de padrão que esses têm o potencial de introduzir nos estudos voltados para a aprendizagem de línguas adicionais. A aplicação de princípios empíricos na área de línguas adicionais está em posição de alterar 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. Uma vez que a investigação da linguagem está calcada em dados, a necessidade de recorrer apenas a falantes que tenham o inglês como língua materna perde o sentido¹³⁴. 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 nativos¹³⁵. Portanto, a LC também é saudada pelos que se

¹³⁴ Reconheço, no entanto, que essa ainda é uma noção forte dentre os estudos em LC. Outrossim, em uma publicação de 1994, Sylviane Granger começaria a questionar essa concepção, ao refletir que não se pode exagerar quanto à importância da adoção de *corpora* de falantes nativos de inglês com vistas ao ensino e aprendizagem dessa língua, porque esses não apontam as questões autênticas atinentes àqueles que necessitam dominar a língua. Segundo ela, faz mais sentido focar em coleções de dados que possam conter erros, se o foco é o ensino de inglês, para que se possa ter noção clara dos pontos fortes e aspectos a desenvolver, em vez de analisar apenas os padrões de uso de usuários reconhecidamente fluentes. Em 2015, o tema foi novamente abordado pela autora em sua plenária (*Learner Corpus Research : A fast-growing interdisciplinary field*) durante a *8th International Corpus Linguistics Conference*, realizada na Universidade de Lancaster/UK, em julho de 2015. A discussão recebeu adesão de autores autores como Douglas Biber, Tony McEnery, John Flowerdew e Stephen Gries, dentre outros. Livro de *abstracts* disponível em <http://ucrel.lancs.ac.uk/cl2015/> (Último acesso em 30/11/2015)

¹³⁵ 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 (RAJAGOPALAN, 2004; WILLIAMS, 2006).

interessam por *EAP* por estar em condições de romper com o paradigma tradicional e fornecer aporte ao trabalho de formação de professores-pesquisadores.

4.1.5. O estatuto do campo

De acordo com Shepherd (2009), “o rótulo Linguística de *Corpus* causa muita comoção” (p. 151) devido à problemática estabelecida em relação à definição do estatuto do campo. Esse dilema já resultou em publicações variadas (SHEPHERD, 2009; MCENERY *et al.*, 2006; TOGNINI-BONELLI, 2001; BERBER SARDINHA, 2000; dentre outros vários). Embora não seja a intenção fechar a questão acerca dos limites do campo, tenciona-se que, com o suporte das caracterizações apresentadas, seja possível perceber a LC pelos múltiplos enquadramentos através dos quais seus estudiosos a cotejam.

Vários de seus principais investigadores, como Berber Sardinha (2000), no Brasil, e McEnery (MCENERY *et al.*, 2006), na Inglaterra, concebem a LC como uma metodologia de pesquisa que possibilita descrever a língua ou testar hipóteses sobre ela. Nesse sentido, há uma interface com teorias de linguagem. Para quem o concebe dessa forma, o campo pode ser combinado a outras abordagens de estudo para a condução de pesquisa¹³⁶. Embora a LC possa ser combinada a abordagens distintas, igualmente, para quem a conceba de outro modo.

Mesmo para os partidários de um enquadramento do campo como uma metodologia de pesquisa, há tendência à cisão dentre os que a percebem por um viés instrumental, no sentido de um conjunto de ferramentas, e os que a concebem como um “modo típico de aplicar um conjunto de pressupostos de caráter teórico” (BERBER SARDINHA, 2000, p. 356).

Kennedy (1998), contudo, reconhece que os praticantes da LC produzem conhecimento novo (p. 9), o que é fato, e, portanto, vai além do que se costuma atribuir a uma metodologia. Entretanto, ele também se refere à LC como fonte de evidências que se originam em *corpora* (KENNEDY, 1999, p. 6). Na reflexão de Sheperd (2009), tal posicionamento, relega o campo ao *status* de pouco mais do que uma fonte de exemplos (p. 152), o que parece

¹³⁶ A exemplo de Adolphs *et al.* (2004) que conjuga LC e Análise da Conversa Etnometodológica (FREITAS e MACHADO, 2008) para qualificar o conhecimento acerca da organização da comunicação em contexto clínico biomédico.

um juízo ainda mais minimizador do que considerá-la uma metodologia.

Outros linguistas de *corpus* (AARTS, 2002; TEUBERT, 2005; WILLIAMS, 2006) se referem ao campo como uma disciplina, o que parece plausível, se, por disciplina, entender-se um *campo de estudos*.

Berber Sardinha (2004), tal qual Hoey (1997), não concebe a LC nem como uma disciplina, nem como uma metodologia, mas como um novo caminho para a Linguística (p. 10), o que nos remete a uma alusão ao campo como desbravador.

Já, Tognini-Bonelli (2001) compreende que a LC “vai muito além de um papel metodológico e se torna uma nova empreitada de pesquisa e uma nova forma de aproximação filosófica para a pesquisa linguística”¹³⁷(p. 1). Juntam-se a ela Leech (1992) e Stubbs (1993). Enquanto Gries (2010) entende que o fato de alguns linguistas de *corpus* defenderem o estatuto da LC como uma teoria se relaciona à forma de investigação com a qual tendem a se identificar, quando pondera o seguinte:

Se os linguistas atribuem o *status* de teoria para a LC ou não, frequentemente, coincide com onde se encontram no *continuum* da linguística que se baseia em *corpus* ou é direcionada por *corpus*. Linguistas que são direcionados pelo *corpus*

- almejam construir teoria sem pressuposição, completamente livres das premissas teóricas pré-*corpus*;
- baseiam as teorizações exclusivamente nos dados do *corpus*;
- costumam rejeitar anotações de *corpus* (uma forma de comprometimento teórico pré-*corpus*) (p. 328)¹³⁸

De parte deste estudo, assume-se este último posicionamento, pelo fascínio de encontrar nos dados e não *nos livros* respostas para os padrões e para as variações sistemáticas que a língua apresenta em relação aos contextos de uso específicos do *corpus* e, a partir daí, construir teoria. Essa possibilidade fornece agência ao pesquisador. Além disso, há aqui uma identificação com a autonomia do caráter interdisciplinar da LC, já que é possível estabelecer cruzamentos com campos de conhecimento variados (como o da LLE ¹³⁹, já que o estudo se

¹³⁷ “*Corpus Linguistics goes well beyond this methodological role and has become a new research enterprise and a new philosophical approach to linguistic enquiry.*”(TOGNINI-BONELLI, 2001, p. 1)

¹³⁸ “*Whether scholars attribute the status of theory to CL or not often somewhat coincides with where they are on the continuum of corpus-driven and corpus-based linguistics. Corpus-driven linguists*

- *aim to build theory from scratch, completely free from pre-corpus theoretical premises*
- *base theories exclusively on corpus data;*
- *often reject corpus annotation (as a pre-corpus theoretical commitment).*” (GRIES, 2010, p. 328.)

¹³⁹ Lembrando uma asserção feita na pág. 12 do cap. 2, a especificidade das linguagens especializadas se expressa principalmente pela frequência de uso de determinados recursos linguísticos comprováveis com o auxílio do método da Linguística Estatística” (HOFFMANN, 2015, p. 41). Ou seja, estamos diante de

ocupa da verificação do *modus dicendi* característico de textos de uma área de conhecimento especializado). É por esses motivos que se confia que a LC representa “uma forma de aproximação filosófica para a pesquisa linguística”, como declara Tognini-Bonelli (2001, p. 1) acima. Afinal, parodiando Fillmore (1992), o embate com os dados está em posição de nos ensinar coisas sobre a linguagem que não descobriríamos de nenhum outro modo (p. 35); o que parece ser um caminho sem volta.

4.2. EAP: o lugar das práticas comunicativas nos contextos acadêmicos

As últimas duas décadas do século XX assistem ao fortalecimento da internacionalização das universidades anglófonas do hemisfério norte. Alia-se à mobilidade acadêmica, a consolidação da língua inglesa na liderança da disseminação do conhecimento científico, de modo que muitos pesquisadores se veem compelidos a adotá-la profissionalmente¹⁴⁰ (CRYSTAL, 2003). Nessa conjuntura, emerge um campo de estudos aplicados que se estabelece com o propósito de descrever a linguagem adotada nos gêneros/registros característicos dos campos acadêmicos e especializados, tanto para produzir materiais pedagógicos, quanto para qualificar o ensino e a aprendizagem da língua (BIBER e CONRAD, 2009, p. 3). Subcampo de *English for Specific Purposes* (ESP), a abordagem denominada *English for Academic Purposes* (EAP)¹⁴¹ se volta àqueles que precisam desenvolver proficiência¹⁴² no uso da língua adicional, assim como àqueles que já demonstram competência sociolinguística em língua inglesa para a consecução de atividades em domínios outros que não os de estudo e pesquisa.

perspectivas muito próximas.

¹⁴⁰ A questão da prevalência da língua inglesa tem ligação estreita com o estabelecimento do campo de EAP. Conforme a perspectiva dos praticantes de pesquisas nesse domínio, o idioma inglês pode ser concebido como língua franca, portanto, facilitadora dos processos comunicativos, ou como um “*Tyrannosaurus rex*, um carnívoro poderoso devorando os outros habitantes das pastagens linguísticas acadêmicas” (SWALES, 1997, p. 374). De minha parte, adoto o vegetarianismo e adiro à primeira posição.

¹⁴¹ Apesar de o campo ser reconhecido a partir da década de 80, o termo *English for Academic Purposes* é atribuído a Tim Jones, que o teria cunhado em 1974 e adotado publicamente em 1977, em uma coleção de artigos acadêmicos editados por Cowie e Heaton (JORDAN, 2002).

¹⁴² Entendida aqui como habilidade para realizar ações sociais específicas através de uma dada língua adicional, em contraste à noção de proficiência como conceito genérico que implica ser capaz de lidar com praticamente todas as realizações sociais com a mesma competência sociolinguística. Para aprofundar o conceito, vide Scaramucci (2000).

Em seus limites de atuação, a área de EAP apresentou-se, inicialmente, voltada para abordagens de estudos. Na atualidade, o campo se preocupa com os processos de letramento acadêmico (LEA e STIERER, 2000). Alguns investigadores da constituição de EAP, como Ken Hyland (2006), apregoam que a abertura de foco representa um movimento em direção a uma perspectiva mais sensível ao contexto, que reflete mudanças tanto na educação superior, quanto no que entendemos por comunicação acadêmica (p. 16).

Assim, não obstante os domínios de EAP abrangerem o desenho de programas de estudos, análise de necessidades de aprendizes e o desenvolvimento de materiais pedagógicos (DUDLEY-EVANS, 2001), o campo comporta um aporte teórico mais inclusivo do que essas designações possam identificar. Na verdade, as demandas acadêmicas contemporâneas, do mesmo modo como as demandas profissionais, não se restringem ao *monitoramento de erros*, como na tradição dos estudos clássicos em aquisição de segunda língua¹⁴³ (ELLIS, 1997). Uma gama de publicações que conjugam EAP, LC e LLE (a exemplo de HOFFMANN, 2015; FADANELLI e FINATTO, 2015; FLOWERDEW, J., 2014; BIBER e GRAY, 2010; HYLAND, 2006; GLEDHILL, 2000; SWALES e FEAK, 2000 e BIBER *et al.*, 1999) atestam a realização de uma profunda revisão de papéis da conta das necessidades de participação efetiva dos usuários em um contexto especializado. A respeito dessa confluência de estudos, os campos que lhes dão origem apresentam estreita ligação.

As pesquisas em EAP se fortalecem com as contribuições da LC, em função, especialmente, de um interesse pela disponibilidade de exemplos de uso autêntico da linguagem. Além disso, a atenção aos padrões e às variações sistemáticas da língua, que representam as investigações em LC, abastecem as descrições de linguagem adotadas nos registros/gêneros característicos dos campos acadêmicos e especializados, que definem o segmento de EAP. Uma grande gama de postulados de Douglas Biber a respeito de gêneros/registros acadêmicos, por exemplo, parte de investigações que correlacionam esses dois campos. Biber e Conrad (2009) relacionam à área “uma aplicação pedagógica aplicada das descrições de registro”¹⁴⁴ (p. 268).

Por extensão, a abordagem de EAP também materializa pedagogicamente os estudos em LLE. Essa interface se define, basicamente, a partir de uma orientação para os discursos de comunicação especializada. Há também uma aproximação compatível no que

¹⁴³ Os estudos clássicos em aquisição de línguas outras que não a materna, nasceram voltados para a aprendizagem de *segunda língua*. O termo era usado para referir a uma língua que falantes, em situação de residência em um país em que esse idioma fosse adotado, precisavam aprender; a exemplo de estudantes em mobilidade acadêmica. (Vide ELLIS, 1997, para conhecer essa temática.)

¹⁴⁴ “... a pedagogical application of register descriptions.” (BIBER e CONRAD, 2009, p. 268)

tange aos 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 científicos. Conforme discutimos anteriormente, Hoffmann (2015) saúda o lugar das descrições dos textos especializados e de seus gêneros, “sobretudo no desenvolvimento da chamada competência textual e na redação de textos autênticos (HOFFMANN, 2015, p. 137). O autor reconhece o papel preponderante que tais descrições desempenham no ensino de línguas.

Face às considerações feitas a respeito da compatibilidade de interesses dentre os três campos, depreende-se que, ao tratarmos de EAP, LC e LLE, estamos diante de perspectivas alinhadas, que se beneficiam mutuamente. Dentre as três áreas, há um foco de atenção que abrange o estudo do uso e dos padrões de uso da linguagem, bem como das condições pragmáticas da comunicação, aspectos eleitos como base analítica para esta pesquisa.

A ênfase de EAP nos letramentos disciplinares, por sua vez, salienta que os discursos e as práticas acadêmicas se realizam de modos distintos nas diferentes áreas de conhecimento. Os textos, suas expectativas de argumentação, as formas de persuasão e de verificação de conhecimento distinguem as áreas de conhecimento umas das outras. A efetividade de um texto acadêmico também se relaciona ao autor ser bem-sucedido em relação a essas diferenças. Por isso tantas pesquisas no campo sublinham as áreas temáticas (como, no caso da saúde, GLEDHILL, 1995; 1999; 2000, 2005, 2012; HYLAND, 2008; CHEN e BAKER, 2010; CORTES, 2004 e LUZON MARCO, 2000, dentre outras que nos servem de base), assim como nosso estudo o faz. Para o usuário, aprender sobre essas distinções implica aprender o conhecimento disciplinar e desenvolver *capital cultural* (BORDIEU, 1991).

Os pesquisadores em EAP também têm se voltado para as distinções dentre os gêneros/registros (HYLAND, 2008; FENG, 2006), bem como dentre os discursos orais e escritos (BIBER e GRAY, 2014; BIBER e GRAY, 2013; CONRAD e BIBER, 2004; HYLAND, 2006). Esses estudiosos demonstram preocupação em comprovar que, assim como os discursos acadêmicos são múltiplos, da mesma forma são os registros que lhes dão forma. Dentre tais pesquisas, chama-se atenção para a de Biber e Gray (2013^a) em que os autores apresentam uma análise léxico-gramatical das diferenças relativas às características das produções escritas e orais de participantes do *TOEFL iBT Test*¹⁴⁵. Nesse caso, os dois

¹⁴⁵ *Test of English as a Foreign Language* em formato *internet-based* (<https://www.ets.org/toefl/ibt/about> Último acesso em 09/12/2015). O TOEFL é, contemporaneamente, o exame de proficiência de inglês como língua adicional de maior adoção internacional.

investigadores verificam a validade do teste como instrumento de aferição de proficiência, com base na variação do discurso dos candidatos, à medida em que se tornam mais proficientes.

No dizer de Hyland,

Mais especificamente, a EAP contemporânea apresenta o objetivo de capturar descrições ‘mais robustas’ de uso da linguagem acadêmica a partir de todas as idades e níveis de proficiência, incorporando e, frequentemente, indo além dos contextos de comunicação imediatos, a fim de entender a natureza do conhecimento disciplinar propriamente dito. O campo faz uso de uma variedade de influências interdisciplinares para seus métodos de pesquisa, teorias e práticas, a fim de fornecer indícios a respeito das estruturas e significado dos discursos acadêmicos orais, escritos, visuais e eletrônicos, com vistas às demandas dos contextos acadêmicos em relação às atitudes comunicativas e em relação às práticas pelas quais essas atitudes podem ser desenvolvidas. EAP consiste, na verdade, em ensino de língua inglesa especializada, com base em demandas sociais, cognitivas e linguísticas das situações acadêmicas alvo, de modo a prover instrução especializada, orientada por uma compreensão de textos e pelas restrições dos contextos acadêmicos.¹⁴⁶ (HYLAND, 2006, p. 2)

Com relação às restrições específicas dos contextos acadêmicos à que Hyland (*op. cit.*) alude acima, os praticantes de EAP identificam três aspectos característicos que influenciam fortemente ensino e pesquisa no campo. São elas:

1. Alta densidade nominal – Na forma de uma proporção significativa de palavras que expressam conteúdo e, portanto, contribuem para tornar a prosa acadêmica mais densa em termos informacionais, em oposição ao uso de palavras gramaticais, tais como preposições, artigos e pronomes. Tomando exemplos de Halliday (1989, p. 3), temos duas sentenças em que a primeira é registrada em estilo acadêmico e a segunda em prosa tipicamente conversacional:
 - a. *‘Investment in a rail facility implies a long-term commitment’.*
 - b. *‘If you invest in a rail facility this implies that you are going to be committed for a long term’.*

¹⁴⁶ “More specifically, current EAP aims at capturing ‘thicker’ descriptions of language use in the academy at all age and proficiency levels, incorporating and often going beyond immediate communicative contexts to understand the nature of disciplinary knowledge itself. It employs a range of interdisciplinary influences for its research methods, theories and practices to provide insights into the structures and meanings of spoken, written, visual and electronic academic texts, into the demands placed by academic contexts on communicative behaviours, and into the pedagogic practices by which these behaviours can be developed. It is, in short, specialized English-language teaching grounded in the social, cognitive and linguistic demands of academic target situations, providing focused instruction informed by an understanding of texts and the constraints of academic contexts.” (HYLAND, 2006, p. 2)

2. Estilo de caráter nominal – Na proporção de ações e eventos apresentados em forma de substantivos, a fim de expressar fenômenos complexos. No exemplo, de Hyland (2006, p. 14) abaixo, identificamos, na segunda sentença, uma transformação dos processos em objetos, em uma prosa caracteristicamente científica:
 - a. *‘The train leaves at 5.00 p.m.’*
 - b. *‘The train’s 5.00 p.m. departure’.*

3. Construções Impessoais – No estilo de uma despreferência pela adoção da primeira pessoa do singular em favor de (a) voz passiva, (b) emprego de sujeitos vazios e (c) uso de retóricos abstratos em que a agência não é atribuída a pessoas, como nos exemplos que se seguem:
 - a. *‘The experiment was designed by in our department.’*
 - b. *‘It was noticed that further reasearch will have to focus on diagnoses.’*
 - c. *‘The data suggest that...’*

(HYLAND, 2006, p. 14)

Ao demonstrarem que a linguagem da ciência é múltipla, variável e, ao mesmo tempo, específica, as pesquisas em EAP marcam o lugar das práticas comunicativas nos contextos acadêmicos. Ou seja, o esforço daqueles que se envolvem com EAP para entender e ensinar os modos como o significado é expresso descortina um caráter multi-letrado dos discursos especializados. Nesse sentido, Hyland (2006) observa que “a comunicação envolve escolhas baseadas em como os textos se organizam em determinados contextos, e os discursos da academia não se baseiam em um único conjunto de regras”¹⁴⁷ (p. 14). Essa acepção desconstrói o mito do déficit, segundo o qual as complexidades do trabalho com os gêneros/registros acadêmicos costumam ser atribuídos exclusivamente a uma falta de proficiência dos usuários. Como se os usuários das práticas acadêmicas *já deversem conhecê-las*, antes mesmo de se defrontarem com seus contextos de uso.

Na verdade, a empreitada parece ser da ordem de revelar que está-se diante de práticas institucionais tão ricas, quanto complexas, mas, ao mesmo tempo, passíveis de sistematização. Não é demérito do usuário, portanto, como marcam posição os estudos em EAP, que essas práticas necessitam ser apre(e)ndidas. Aspectos esses que Bakhtin (1979),

¹⁴⁷ “...communication involves making choices based on the ways texts work in specific contexts and that the discourses of the academy are not based on a single set of rules.” (HYLAND, 2006, p. 14)

Swales e Feak (2009) além de Biber e Conrad (2009), tão bem defendem, acerca dos gêneros secundários, no caso do primeiro, e acerca das narrativas acadêmicas, nos casos dos demais.

É, igualmente, para marcar o lugar das práticas comunicativas nos contextos acadêmicos, conforme mencionado acima, que esta tese se coloca. Com a realização do trabalho, almeja-se contribuir com as investidas para qualificar os programas brasileiros de ensino de EAP, bem como para a formação de educadores linguísticos que atuam nesse campo. Conseqüentemente, mira-se na formação de usuários mais informados acerca das diversidades das linguagens acadêmicas. O ponto de partida são as questões probabilísticas de adoção de padrões lexicais e contexto de uso, no cruzamento entre EAP e Ciências da Saúde.

4.3. Desempacotando o léxico dos textos

A agenda da descrição da linguagem pelo ponto de vista do léxico é a identificação dos agrupamentos lexicais mais frequentemente atestados pelo uso. Essa meta parte do princípio idiomático, postulado por Sinclair (1987), de acordo com o qual “um usuário da língua tem a sua disposição um grande número de frases semi-pré-construídas, que constituem escolhas únicas, mesmo que elas pareçam ser analisáveis em segmentos”¹⁴⁸ (p. 120).

Dutra e Berber Sardinha (2015) asseveram que é recente o estabelecimento da Fraseologia¹⁴⁹, como uma disciplina e não apenas como um campo da Lexicologia (p. 60). Os autores complementam, com base em Sinclair (2008), que “o ponto de partida dos estudos da fraseologia é a frase por ser ela a unidade primária do significado” (p. 60). São os avanços da LC, conjugados à adoção de *corpora* eletrônicos, que têm permitido essa especialização. De acordo com Gries (2008, APUD, DUTRA e BERBER SARDINHA, *op. cit.*), os avanços da tecnologia têm possibilitado a adoção de um ferramental, tal qual a geração de listas de frequência de *n*-gramas ou linhas de concordância, que possibilita a identificação das combinações de palavras recorrentes nos discursos.

A visão de que a linguagem é formada por porções lexicais tem dado margem a um número de estudos, organizados desde longa data, por perspectivas e contextos diversos,

¹⁴⁸ “The language user has available to him a large number of pre-constructed or semi-pre-constructed phrases that constitute single choices, even though they appear to be analysable into segments” (SINCLAIR, 1987, p.120).

¹⁴⁹ Entendida aqui como “o estudo da estrutura, significado e uso das combinações de palavras”. (COWIE, 1998, p. 3168)

que adotam denominações igualmente diferenciadas para seu objeto de análise (JESPERSEN, 1924; PALMER, 1933; ALTENBERG, 1993; KJELLMER, 1991; MOON, 1998; NATTINGER e DECARRICO, 1992; SINCLAIR e RENOUF, 1988, BIBER *et al.*, 1999, para citar alguns).¹⁵⁰ Um dos primeiros linguistas a chamar atenção para as combinações recorrentes de palavras foi Firth (1957), que, como mencionado, denominou-as de colocações (*collocations*). Contudo, Firth não ofereceu uma definição explícita para o fenômeno. Um pouco mais tarde, Halliday *et al.* (1964) apresentou uma definição que enfatizou “a tendência de um item lexical a co-ocorrer com uma ou mais palavras”¹⁵¹ (p. 33).

As combinações de palavras podem incluir, *grosso modo*, dois tipos de blocos. Há os blocos de caráter relativamente fixo, que são passíveis de reconhecimento pela intuição dos linguistas e que, geralmente, se apresentam estruturalmente completos, a exemplo de ‘*kick the bucket*’ e ‘*eat one’s cake and have it*’. Há, contudo, os blocos de palavras cujos componentes são fixos em relação à recuperação, mas podem variar, em alguma medida, em relação à forma estrutural. Esses são passíveis de reconhecimento pela frequência de ocorrência e se apresentam, na maioria das vezes, como estruturalmente incompletos, a exemplo de ‘*in order to*’ e ‘*if it’s possible to*’.

A terminologia de referência para as combinações de palavras do primeiro bloco abrange, dentre as denominações mais frequentes, *pre-fabs* (BOLINGER, 1976; GRANGER, 1998), *phrasicons* (DE COCK *et al.*, 1998), *lexemas frasais* (MOON, 1998), *enquadramentos colocacionais* (RENOUF e SINCLAIR, 1991), *colocações* ou *idioms* (SINCLAIR, 1991), *fraseologias* (em inglês, *phraseology*) (BLAIS, 1993; BEVILACQUA, 2004; GRANGER e MEUNIER, 2008; MEUNIER e GRANGER, 2007) e *linguagem ou sequências formulaicas* (SCHMITT, 2004; WRAY, 2000, 2008). No caso desse tipo de bloco, sobra pouca ou mesmo nenhuma escolha linguística variacional para o usuário.

Já os grupos poli-lexicais, aqueles do segundo bloco, são identificáveis através de programas computacionais extratores. Esses são recuperados através de um trabalho direcionado por *corpus* (*corpus-driven*), considerando um critério de frequência e distribuição ao longo do acervo de dados. A terminologia de referência desses blocos de palavras envolve, dentre outros, *clusters* (SCOTT, 1996), *unidades de multi-palavras* (ROBINSON e ELLIS,

¹⁵⁰ Não obstante o interesse pelos padrões recorrentes da linguagem, Conrad e Biber (2004) criticam o fato de que a Lexicografia mantém ênfase na palavra individual como base do discurso, para organização de dicionários. Para os autores, tal característica parece perpetuar que, para esse campo, as frases e orações se organizam a partir de unidades individuais e não a partir de blocos de palavras associadas (p. 56).

¹⁵¹ “...the tendency of a lexical item to co-occur with one or more other words.” (HALLIDAY *et al.*, 1964, p. 33).

2008), *pacotes lexicais* (em inglês, *lexical bundles*), também reconhecidos como *feixes lexicais* (BIBER *et al.*, 1999; BIBER e CONRAD, 1999; BERBER SARDINHA, 2000, 2003; BIBER, CONRAD e CORTES, 2003, 2004; BIBER *et al.*, 2004; CORTES, 2002, 2004, 2006, 2008; CONRAD e BIBER, 2004; BIBER, 2006; STUBBS, 2007; SCOTT e TRIBBLE, 2006; HYLAND, 2008a) ou *n-gramas* (SINCLAIR, 2004).

Chama-se atenção para o aspecto de que, de acordo com a análise de Scott e Tribble (2006, p. 131), essa última categoria de associações de palavras é mero resultado artificial a partir de programas extratores, com base em critérios distributivos. Ou seja, o critério para sua identificação é a recorrência em um acervo de dados, a partir de um número pré-estabelecido de vezes.

Muitos estudiosos do campo compartilham o entendimento de que a fluência está relacionada ao uso das associações de palavras tipicamente constantes. De acordo com Hyland (2008^a), o uso extensivo de sequências pré-fabricadas, a exemplo de *'as it has been noted'*, em gêneros acadêmicos, ajuda a sinalizar o registro para os leitores, assim como reduz o tempo de processamento dos textos pela adoção de padrões familiares para unir elementos, contendo informações novas (p. 5).

Este trabalho se alinha a essa concepção que relaciona fluidez ao uso de combinações recorrentes de palavras, porque defende-se que é a frequência de ocorrências que 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). Ou seja, sabe-se 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. São, entretanto, 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 nosso interesse demonstrar em relação ao acervo reunido. Para tanto, cada elemento ou item lexical do discurso 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). É, portanto, em função de o estudo do uso e dos padrões de uso, bem como das condições pragmáticas da comunicação, terem sido eleitos como base analítica para a investigação que a pesquisa explora o tema dos agrupamentos de palavras recorrentes. Assumido esse posicionamento, restava então definir que perspectiva e denominação adotar como forma de aproximação dos blocos repetidos que integram o repertório dos textos acadêmicos.

4.3.1. Da escolha por pacotes lexicais (ou *lexical bundles*)

Optou-se pela segunda categoria de grupos poli-lexicais, aqueles cujos componentes podem variar, por serem mais desafiantes para o aprendiz, uma vez que funcionam como blocos de construção para a linguagem acadêmica.¹⁵² Além do mais, esses blocos de palavras co-ocorrentes são relevados por sua frequência, através de estudos direcionados por *corpus*. Outrossim, em termos de ponto de partida epistêmico, corrobora-se a perspectiva variacionista, que implica reconhecer a influência sociocultural dos usuários e acatar o caráter plástico dos discursos (BIBER *et al.*, 1999). Aceita-se ainda que, sendo os indivíduos expostos a diferentes estruturas e padrões linguísticos, é necessário reconhecê-los, interpretá-los e reproduzi-los, para promover o uso efetivo, aos moldes dos estudos em EAP (BIBER e CONRAD, 2009; HYLAND, 2008^a). Ademais, valora-se o foco em sistematicidade que se realiza pela frequência de ocorrências de itens lexicais, que tipificam os registros da comunicação (BIBER, 1995; 2006). Em outras palavras, a partir dessas confluências, naturalmente, culmina-se por eleger a denominação consagrada pelos estudiosos que inspiraram a organização da pesquisa (BIBER *et al.*, 1999; BIBER e CONRAD, 1999; BERBER SARDINHA, 2000, 2003; BIBER, CONRAD e CORTES, 2003, 2004; BIBER *et al.*, 2004; CORTES, 2002, 2004, 2006, 2008; CONRAD e BIBER, 2004; BIBER, 2006; STUBBS, 2007; SCOTT e TRIBBLE, 2006; HYLAND, 2008^a, 2008^b, 2008^c).

Para definir as associações de palavras mais frequentes que integram o repertório dos usuários de uma língua, destacam-se os termos do primeiro estudioso a nomeá-las como *pacotes lexicais*: “combinações de três ou mais palavras que recorrem mais frequentemente em um dado registro, independentemente de sua idiomaticidade e de seu *status* estrutural. Ou seja, pacotes lexicais são simplesmente sequências de formas de palavras que comumente se associam em discurso natural”¹⁵³ (BIBER *et al.*, 1999, p.990); a exemplo de ‘*on the other hand*’, ‘*in the present study*’ e ‘*are likely to be*’. Mais tarde, Biber se referiria

¹⁵² Como menciona Hyland (2008^a), enquanto aprendizes da língua inglesa se esforçam para aprender *phrasal verbs* como ‘*look after*’ ou ‘*idioms*’ como ‘*beat around the bush*’, esses são relativamente raros, se comparados às sequências de palavras recorrentes identificáveis pela alta incidência (p. 4).

¹⁵³ “*Combinations of words that in fact recur most commonly in a given register. Lexical bundles are recurrent expressions, regardless of their idiomaticity and regardless of their structural status. That is, lexical bundles are simply sequences of words forms that commonly go together in natural discourse.*” (BIBER *et al.*, 1999, p. 990)

a *lexical bundles* simplesmente como “as sequências lexicais mais frequentes encontradas em um registro”¹⁵⁴ (BIBER *et al.*, 2004, p. 376).

4.3.2. Das principais características dos pacotes lexicais

De acordo com Conrad e Biber (2004), “até mesmo pesquisadores que concordam com a importância das sequências multi-palavras, discordam acerca dos métodos de identificá-las e estudá-las”¹⁵⁵ (p. 57). Os dois investigadores, por sua vez, compreendem que os critérios a serem adotados para reconhecê-las estão fortemente atrelados aos propósitos de estudo desses agrupamentos de palavras. Praticantes de investigações orientadas por *corpora*, Conrad e Biber (*op. cit.*) defendem que descrever associações entre palavras requer primeiramente a identificação dos diferentes tipos de grupos poli-lexicais encontrados em uma coleção de dados que tipificam um registro. A empreitada, em outras palavras, é distinta de catalogar expressões recorrentes de modo meramente apriorístico, simplesmente porque *seria didático ensiná-las desse modo*.

Para Conrad e Biber (*op. cit.*), a perspectiva com referência ao estudo desse fenômeno de linguagem é da seguinte ordem:

Nossas perguntas de pesquisa em relação a esta abordagem são exploratórias. Queremos saber se existem sequências multi-palavras empregadas com alta frequência nos textos, se registros diferentes tendem a usar diferentes tipos dessas sequências e, se sim, com que extensão os pacotes lexicais desempenham funções discursivas e, portanto, desempenham uma parte importante no repertório comunicativo de falantes e escritores.¹⁵⁶ (CONRAD e BIBER, 2004, p. 58.)

Constata-se, portanto, que a abordagem de identificação que os dois autores concebem é de ordem empírica e não meramente intuitiva.

Ao longo do detalhado estudo que organizou uma gramática da língua inglesa com base em trabalho de *corpus*, intitulada *Longman Grammar of Spoken and Written*

¹⁵⁴ “The most frequent recurring lexical sequences in a register.” (BIBER, *et al.*, 2004, p. 376)

¹⁵⁵ “Even scholars who agree on the importance of multi-word sequences often disagree about the methods for identifying and studying them” (CONRAD e BIBER, 2004, p. 56)

¹⁵⁶ “Our research questions in this approach are exploratory. We ask whether there are multi-word sequences that are used with high frequency in texts, whether different registers tend to use different sets of these sequences, and, if so, to what extent the bundles fulfill discourse functions and thus play an important part in the communicative repertoire of speakers and writers.” (CONRAD e BIBER, 2004, p. 58.)

*English*¹⁵⁷ (BIBER *et al.*, 1999), Biber não apenas nomeia as principais associações recorrentes de palavras que animam a prosa acadêmica e a conversa cotidiana, mas igualmente relaciona o item frequência como a característica mais saliente desses feixes lexicais. A particularidade da constância de ocorrência se mostra fundamental, uma vez que determina se um grupo de palavras constitui ou não um pacote lexical.

O ponto de corte de frequência, entretanto, é estabelecido de modo arbitrário (BIBER *et al.*, 2004, p. 376), embora Biber *et al.* (1999) especifique que os pacotes lexicais estariam restritos a combinações recorrentes verificáveis pelo menos 10 vezes por milhão de palavras. Muitos estudos do tema, entretanto, se organizam a partir de *corpora* menores, o que exige ajustes nesse padrão. Outro critério para a averiguação de pacotes lexicais é a dispersão. De acordo com esse quesito, é consenso (BIBER *et al.*, 1999; BIBER e BARBIERE 2007 e CORTES, 2004) que uma sequência de palavras deve acontecer em, pelo menos, de 3 a 5 textos diferentes de um *corpus*, para contar como um agrupamento poli-lexical e, com isso, salvaguardar a possibilidade de não representar apenas uma idiosincrasia do estilo do autor.

Biber *et al.* (1999) e Hyland (2008^a) concordam que pacotes lexicais podem ser entendidos como colocações estendidas, no sentido de que sequências de três ou mais palavras demonstram uma tendência estatística a co-ocorrer em um registro. Essencialmente, estamos tratando de palavras que se apresentam de modo sequencial mais comumente do que seria esperado sob a mera ocorrência do acaso e que trabalham para dar forma ao significado em contextos específicos, assim como contribuem para a coerência de um texto (HYLAND, *op. cit.*, p. 4-5).

Cortes (2004) descreve os pacotes lexicais como fixos, já que os programas computacionais que os recuperam nos textos estabelecem buscas por formas estáveis (p. 400). No entanto, esse caráter fixo é definido exclusivamente com base em frequência. No *corpus* dessa pesquisadora, por exemplo evidenciou-se como pacote lexical ‘*these results suggest that*’, entretanto, a forma singular dessa sequência, ‘*this result suggests that*’, não se apresentou como colocação estendida, tendo em vista a linha de corte previamente estabelecida para número de ocorrências. Por esse motivo, sustenta-se acima que os pacotes

¹⁵⁷ Esta é a primeira gramática da língua inglesa que dedica uma seção ao tema dos pacotes lexicais. Além disso, a obra é seminal não só porque se organizou a partir de *corpus*, mas também por adotar a perspectiva de registro, bem como, pelo tamanho imenso da coleção de dados cotejados (5 milhões de palavras para cada registro), ainda porque baseou-se exclusivamente em frequência e focou-se em agrupamentos de palavras mais longos do que estudos anteriores abordaram (sequências de 4, 5 e 6 palavras).

lexicais se constituem em blocos fixos quanto à recuperação, mas, por vezes, variáveis em relação à forma estrutural.

De acordo com Biber *et al.* (2004), em se tratando de prosa acadêmica, as sequências de palavras mais frequentes de um registro” (p. 371)¹⁵⁸ costumam estabelecer pontes entre frases, que podem ser de ordem nominal (como em ‘*a wide range of*’) ou que podem relacionar um nome e um agrupamento verbal (como em ‘*likely to be considered*’). Os pacotes lexicais típicos dos registros acadêmicos são formados, mais frequentemente, por elementos nominais e preposicionais (a exemplo de ‘*in the process of*’ e ‘*a part of the*’). Já em registros do tipo conversa cotidiana, apresentam-se como segmentos frasais, na forma de estruturas declarativas (a exemplo de ‘*I don’t know what*’ e ‘*I told him*’) ou mesmo na forma de estruturas interrogativas (como em ‘*Can I have a...?*’ e ‘*What are you doing?*’).

Comumente, apresentam-se através de formas estruturais incompletas (como em ‘*I don’t know if*’ e ‘*the nature of the*’), portanto, não constituem unidades que os linguistas tendem a reconhecer por intuição¹⁵⁹. A esse respeito, vale ressaltar, como nota Biber *et al.* (1999), que a recorrência de pacotes lexicais em forma de unidades com estrutura gramatical incompleta, obscurece a distinção dentre léxico e gramática, que vem a ser um dos princípios centrais do programa de pesquisa de Sinclair.

Por outro lado, Biber *et al.* (1999) assevera que, “quando um pacote lexical é estruturalmente completo em prosa acadêmica, apresenta-se tipicamente na forma de um grupo preposicional que funciona como sinalizador do discurso”¹⁶⁰ (p. 999) (a exemplo de ‘*in the present study*’, ‘*in the same way*’ ou ‘*in future studies*’). De todo modo, Biber (*op. cit.*) constata que apenas 5% dos pacotes lexicais encontrados na prosa acadêmica representam unidades estruturais completas.

Outro aspecto adotado para caracterizar as sequências multi-palavras é o fato de que, no geral, costumam ser semanticamente transparentes. Esse ponto os diferencia, por conseguinte, das expressões idiomáticas ou fixas (CORTES, 2004, p. 400). Não obstante, Dutra e Berber Sardinha (2015) complementam que

Se, por um lado, há pacotes lexicais como ‘*on the other hand*’ que tem significado de contraste claramente detectável em uma análise isolada do pacote, há outros como ‘*in which the*’ que exigem uma observação do contexto linguístico de ocorrência para determinar, por exemplo, que esse pacote pode ser considerado uma

¹⁵⁸ “*The most frequent sequences of words in a register.*” (BIBER, CONRAD e CORTES, 2004, p. 371)

¹⁵⁹ Entretanto, apresentam fortes correlatos gramaticais, o que facilita sua identificação, como será apresentado na próxima subseção com base na taxonomia estrutural de Biber *et al.* (1999).

¹⁶⁰ “*When a lexical bundle is structurally complete in academic prose, it is typically a prepositional phrase that functions as a discourse signalling device ...*” (BIBER *et al.*, 1999, p. 999)

expressão referencial de enquadramento de atributo intangível¹⁶¹. (DUTRA e BERBER SARDINHA, 2015, p. 61)

A esse respeito, Conrad e Biber (2004) notam que grupos comuns de sequências relativamente fixas de palavras apresentam funções discursivas identificáveis nos textos. É por esse motivo que sequências de apenas duas palavras não são consideradas como pacotes lexicais, uma vez que tendem a não apresentar funções discursivas identificáveis, sendo consideradas fragmentos de pacotes maiores. Já os agrupamentos lexicais de três palavras gráficas costumam ocorrer com maior frequência por serem do tipo estendido. Isso significa que, se extrairmos de uma dado *corpus* um pacote lexical composto de um conjunto de três palavras, a exemplo de ‘*the nature of*’, provavelmente, encontremos ‘*the nature of this*’ ou ‘*the nature of this paper*’, como sequência de encadeamento. É por isso que Biber *et al.* (1999, p. 992) destaca que pacotes lexicais que contêm 5 ou 6 palavras, ou seja, ou mais longos, são mais fraseológicos, em virtude de encapsularem agrupamentos menores em sua formação. É também por esse motivo que grupos poli-lexicais mais longos apresentam menos probabilidade de ocorrer e, portanto, o ponto de corte pode variar, conforme a extensão do pacote definida pelo pesquisador ou conforme o tamanho do *corpus* disponível.

Em síntese, as caracterizações descritas sumarizam os seguintes aspectos acerca de pacotes lexicais:

1. Representam as sequências de palavras mais frequentes de um registro;
2. Apresentam-se como associações de mais de duas palavras;
3. São identificáveis empiricamente pela frequência de ocorrência e dispersão ao longo de um *corpus*;
4. Constituem-se em colocações estendidas pela tendência à recorrência em um dado registro;
5. São detectáveis através de aspectos estabelecidos de modo arbitrário;
6. Podem apresentar variação estrutural, mas são extraídos em formas fixas;

¹⁶¹ De acordo com Dutra e Berber Sardinha (2015), o termo *atributo intangível* comporta um escopo mais amplo do que porções (p. 61), tendo sido cunhado para definir expressões que são “simplesmente sequências de palavras que comumente co-ocorrem em discurso natural.” (BIBER *et al.*, 1999, p. 990)

7. Costumam variar de um registro para outro;
8. São, na maioria das vezes, 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'*);
9. Tendem a ser mais fraseológicos, à medida que a sequência de encadeamento se apresenta mais longa;
10. Apresentam fortes correlatos gramaticais, dividindo-se em tipos estruturais básicos ao longo de um *corpus*;
11. Costumam estabelecer ligações entre frases de caráter nominal ou entre um nome e um agrupamento verbal em prosa acadêmica, assim como são, mais frequentemente formados por elementos nominais e preposicionais;
12. Tendem a ser frasais e a se apresentarem em forma de estruturas declarativas ou interrogativas em conversa cotidiana;
13. Na maioria das vezes, são semanticamente transparentes, em oposição a expressões de caráter idiomático, embora a compreensão possa exigir observação do contexto linguístico de ocorrência.

4.3.3. Das classificações dos pacotes lexicais

Para entender o uso desses agrupamentos recorrentes de palavras, é necessário conhecer tanto a composição lexical, quanto os objetivos funcionais que exercem nos textos de um dado gênero/registo. Sobre esse segundo objetivo, Biber e Barbieri (2007) apregoam que os pacotes lexicais trabalham como “um tipo de ‘cabeça’ pragmática de frases mais

longas ou orações, em que funcionam como marcadores discursivos para expressão de informações novas”¹⁶² (p. 270).

Por isso, o trabalho de análise dos pacotes lexicais costuma se orientar por dois eixos:¹⁶³

- (1) Estruturalmente, enfocando a composição lexical, estabelecida a partir dos padrões mais usuais encontrados para agrupamentos de três ou mais palavras gráficas e/ou
- (2) Funcionalmente, em relação à estrutura retórica dos textos (BIBER *et al.*, 1999, pp. 997–1025; BIBER *et al.*, 2004; BIBER, 2006; HYLAND, 2008^a).

A classificação estrutural da *Longman Grammar of Spoken and Written English* (BIBER *et al.*, 1999) tem sido amplamente adotada para estudos acerca de combinações recorrentes de palavras (CORTES, 2004; HYLAND, 2008^a), independentemente de os pacotes lexicais não representarem unidades estruturais completas. Três categorias estruturais amplas, que se agrupam em várias outras subcategorias, são distinguidas:

1. NP (*Noun Phrases*)^{164 165} - baseiam-se em frases nominais, com fragmentos pós-modificadores, a exemplo de ‘*the role of the*’ e ‘*the way in which*’;
2. PP (*Prepositional Phrases*) – baseiam-se em uma preposição seguida de uma frase nominal, a exemplo de ‘*at the end of*’ e ‘*in relation to the*’;
3. VP (*Verbal Phrases*) – baseiam-se em qualquer combinação de palavras com um componente verbal, a exemplo de ‘*in order to make*’ e ‘*was one of the*’.

A tabela que se segue exemplifica as categorias e subcategorias descritas:

¹⁶² “...a kind of pragmatic ‘head’ for larger phrases or clauses, where they function as discourse frames for the expression of new information”. (BIBER e BARBIERE, 2007, p. 270).

¹⁶³ O registro de ambas as taxonomias, bem como de respectivos exemplos, tem objetivo didático de apresentar o que a literatura descreve como ponto de partida para as análises estruturais e funcionais. Não se está inferindo aqui que cada item dessas classificações estará presente em todo *corpus*, tampouco que todo *corpus deva necessariamente se encaixar nessa* descrição.

¹⁶⁴ Uma vez que o trabalho analítico deste estudo se desenvolve a partir de *corpus* em língua inglesa, por questões de praticidade analítica e guardadas as diferenças entre padrões léxico-gramaticais característicos, optei por manter as categorias no original em inglês e detalhá-las em português.

¹⁶⁵ Cortes (2005) averiguou que 60% dos pacotes lexicais da prosa acadêmica se inserem nesta categoria.

Tabela (2): Classificação Estrutural de Pacotes Lexicais de acordo com Biber *et al.* (1999)

	Padrões Estruturais:	Exemplos:
i.	<i>Noun phrase with an of-phrase fragment</i>	<i>a function of the, a wide range of, aim of this study, one of the most, the aim of this, the structure of the...</i>
ii.	<i>Prepositional phrase with embedded of-phrase fragment</i>	<i>as a function of, for the production of, in the case of, in the absence of, in the development of, in the presence of..</i>
iii.	<i>Other prepositional phrases (fragment)</i>	<i>in this paper we, in this study we, in this work we, of this study was, on the other hand, with respect to the, ...</i>
iv.	<i>Passive verb + prepositional phrase fragment</i>	<i>be used as a, can be used as, is based on the, is related to the, is seen as a, considered as a, ...</i>
v.	<i>((Noun +) Verb phrase +) that-clause fragment</i>	<i>results suggest that the, the results show that, we show that the, we find that the, we consider that the, ...</i>
vi.	<i>(Verb phrase +) to-clause fragment</i>	<i>can be used to, have been developed to, this work aims to, to be used as, to be used as, was found to be, ...</i>
vii.	<i>Pronoun/noun phrase + be (+...)</i>	<i>there is not a, this paper is to, this study was to, this work was to,...</i>
viii.	<i>Anticipatory it patterns (it + verb/adjective phrase)</i>	<i>it is shown that, it is possible to, it was possible to,</i>
xix.	<i>Copula be + noun phrase</i>	<i>is one of the, is the result of, ...</i>
x.	<i>Other verb phrases</i>	<i>here we report the, paper we propose a, this paper presents the, this paper we present, this paper we propose, ...</i>
xi.	<i>Other expressions</i>	<i>as well as the, as wells as to, in order to get, ...</i>

Fonte: Freitas (2016)

Os pacotes lexicais também foram classificados por Biber *et al.* (2004, p. 384) com base em sua funcionalidade ou nos objetivos pragmáticos. Nesse caso, as funções taxonômicas se referem ao significado e aos 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. Estabelecem-se três macro categorias analíticas.

Tabela (3): Classificação Pragmática de Pacotes Lexicais de acordo com Biber *et al.* (1999)

Padrões Indicadores de Postura Epistêmica (Stance Bundles):	Exemplos:
Estado Epistêmico (ou avaliação do autor sobre uma proposição em termos de certeza ou incerteza)	<i>I don't know, I don't think so, ...</i>
Atitude do Autor (acerca de uma proposição, obrigação ou diretiva)	<i>it's important to, it's necessary to, that need to be, it has to be, ...</i>
Habilidade (do autor para fazer algo)	<i>will be able to, it's difficult to do, to be able to, ...</i>
Introduzir um Tópico	<i>I want to talk about, This paper addresses a...</i>
Elaborar acerca do Tópico	<i>be taken into account, in more detail in, on the other hand, on the other hand, can be used to, ...</i>
Fazer Inferências	<i>in the sense that, as a result of, in view of the, this is due to, ...</i>
Padrões Indicadores Referenciais (Referential Bundles)	Exemplos
Pacotes Lexicais de Enquadramento especificam um dado atributo ou condição	<i>in the context of, in terms of the, the nature of, the existence of a, ...</i>
Expressões de Quantificação qualificam uma proposição com expressões relacionadas ao que seja potencialmente mensurável, como tamanho, número, quantidade ou extensão	<i>per cent of the, a wide range of, the extent to which, in a number of, ...</i>
Pacotes Lexicais de Lugar/ Tempo ou Dêiticos	<i>at the beginning of, at the same time</i>

Fonte: Freitas (2016)

Saliente-se que essa taxonomia é passível de adaptação ou modificação, com base nos traços característicos dos diferentes estudos, como o fizeram Stubbs (2002), Biber *et al.* (2004), Scott e Tribble (2006), Cortes (2006, 2008) e Hyland (2008^a), dentre outros que adotaram metodologias de análise diferentes, para os aspectos estruturais e pragmáticos.

Do mesmo modo, nos termos de Dutra e Berber Sardinha (2015), “a inclusão de pacotes em certas categorias não indica que eles não possam ter outra função no discurso” (p. 58). Ou seja,

Essas categorias funcionais – determinadas após estudo das frases em contexto usando-se um concordanciador – não devem ser consideradas como definitivas ou exclusivas, pois muitas fórmulas têm múltiplas funções, e sim como indicação da função mais saliente que as frases exercem em contextos acadêmicos. (SIMPSON-VLACH e ELLIS, 2010, APUD, DUTRA e SARDINHA, 2015, p. 58)

Não obstante, o fato dos pacotes lexicais demonstrarem características estruturais e funcionais identificáveis, como já demonstrado em um número de estudos exploratórios (BIBER e CONRAD, 1999, BIBER *et al.*, 2003, BIBER *et al.*, 2004, Hyland 2008^a, 2008^b, 2008^c), torna-os um bom ponto de partida para a exploração das diferenças entre os padrões colocacionais dos gêneros/registros, disciplinas e grupos de autores (RÖMER, 2009).

4.3.4. Das perspectivas das investigações

Um dos primeiros estudos acerca do tema de pacotes lexicais descritos na literatura é o de Altenberg (1993, 1998) que, a partir do London-Lund Corpus, analisou “combinações de palavras recorrentes”¹⁶⁶ (p.113) em língua inglesa e demonstrou que essas são “evidentes em todos os níveis de organização linguística”¹⁶⁷. Esse autor constatou, outrossim, que cerca de 80% das palavras em um corpus formam partes de combinações recorrentes de palavras.

Os estudos de Biber e seus colegas (BIBER e GRAY, 2013; BIBER e BARBIERE, 2007; BIBER, CONRAD e CORTES, 2003, 2004; BIBER, JOHANSSON,

¹⁶⁶ “*recurring word-combinations*” (ALTENBERG, 1998, p. 113).

¹⁶⁷ “*evident at all levels of linguistic organization*” (ALTENBERG, 1998, p. 113).

LEECH, CONRAD e FINNEGAN, 1999; BIBER e CONRAD, 1999) têm tido por objetivo cotejar as distinções entre registros em termos do padrão dos pacotes lexicais. Através desses, confirmou-se que a conversa cotidiana e a prosa acadêmica apresentam padrões de distribuição de grupos poli-lexicais distintos. Por exemplo, a maioria dos agrupamentos recorrentes de palavras presentes na conversa cotidiana são oracionais (como em *'I don't know'* e *'I thought it was'*), enquanto que, na prosa acadêmica, os pacotes lexicais costumam ser frasais (como em *'a wide range of '* e *'in this paper we'*). As correlações léxico-gramaticais, desse modo, facilitam os agrupamentos em tipos estruturais básicos. Com esses estudos, Biber e os autores a ele associados (*op. cit.*) concluem que os pacotes lexicais, na verdade, representam a função de blocos básicos de construção do discurso.

Uma quantidade de outras pesquisas, dentre as quais este estudo se insere, têm focalizado o papel da proficiência e familiaridade dos escritores com a comunidade a que pertencem e o quanto a competência no uso de uma língua pode afetar a presença de pacotes lexicais. Viviana Cortes (2002), por exemplo, verificou que os pacotes lexicais produzidos por estudantes universitários se mostravam funcionalmente diferentes daqueles produzidos por autores de publicações. Cortes (*op. cit.*) infere que a função principal das associações recorrentes de palavras é auxiliar os estudantes a escreverem e falarem com maior fluência. Em estudo de 2004, a autora constaria que noviços em escrita acadêmica raramente utilizavam os grupos poli-lexicais adotados pelos pesquisadores com bagagem de publicação.

Já Hyland (2008^a) constatou a ocorrência de variação de pacotes lexicais de quatro palavras, ao examinar um corpus de 3.5 milhão de palavras de quatro disciplinas diferentes em termos de forma, estrutura e função. Em outro estudo (HYLAND 2008^b), em que esse pesquisador contrastou produções de pós-graduandos a publicações de artigos, o autor constatou que pesquisadores menos experientes tendem a se basear mais em linguagem formulaica. Nesse caso, Hyland (*op. cit.*) analisa que a adoção de um maior número de associações recorrentes de palavras é um modo de os noviços reforçarem seu pertencimento à comunidade acadêmica. Um outra conclusão relevante de Hyland (2008^a, 2008^b) é de que há variação na quantidade de uso de pacotes lexicais dentre os diferentes campos de conhecimento. O campo da Engenharia, comparado à Biologia, Administração de Empresas e Linguística Aplicada, por exemplo, caracterizou-se como o que apresenta mais alta incidência de uso de agrupamentos recorrentes de palavras.

Em um estudo de 2010, Salazar analisou pacotes lexicais em duas variedades de inglês (filipino e britânico), através de uma análise de combinações recorrentes de palavras resgatadas de dois *corpora* de publicações de artigos de pesquisa, em Medicina. Os resultados

quantitativos apontaram que o *corpus* filipino continha um número menor de pacotes lexicais do que o britânico, enquanto os resultados qualitativos revelaram diferenças estruturais e funcionais entre os dois *corpora*. Essas constatações sugerem distinções no uso grupos poli-lexicais inerentes às variedades da língua inglesa.

Destaca-se também o trabalho de Salazar (2011), um estudo de *corpus* sobre pacotes lexicais em inglês científico, investigação em que a autora se debruça sobre frequência, estrutura e função de colocações estendidas. Agrupamentos de 3 a 6 palavras foram extraídos de um corpus de 1.3 milhão de palavras originárias de artigos publicados dos campos da Biologia e Bioquímica. A autora relaciona o *corpus* mencionado a um outro menor, com produções de cientistas espanhóis. A comparação revelou que estes últimos fazem uso excessivo de certos pacotes lexicais, o que configura repetitividade e falta de variação.

Chen e Baker (2010), por sua vez, concluem que o uso de pacotes lexicais varia tanto em termos de frequência, quanto de aspectos pragmáticos, na escrita de autores profissionais em comparação a estudantes novíços. Esses pesquisadores evidenciaram que os autores experientes adotaram significativamente mais pacotes lexicais do que os aprendizes, além de apresentarem padrões funcionais diferentes, o que vai ao encontro das evidências de Salazar (2010) e de encontro às constatações de De Cock (2000), Hyland (2008^a, 2008^b, 2008^c) e Salazar (2011). Portanto, tais conclusões parecem apontar para necessidade de mais estudos no tema.

Por fim, com o objetivo de melhor compreender a escrita de redações em inglês de alunos universitários brasileiros, Dutra e Berber Sardinha (2015) analisaram 200 pacotes lexicais de maior frequências com base no Br-ICLE¹⁶⁸ (*corpus* em formação com redações de alunos brasileiros de diversas universidades do país), no ICLE (*International Corpus of Learner English*) e no LOCNESS (*Louvain Corpus of Native English Essays*). O estudo revelou alta frequência de expressões referenciais, aspecto característico dos gêneros textuais acadêmicos, assim como diferenças significativas dentre as categorias pragmáticas analisadas. O Br-ICLE apresentou, por exemplo, pouca variedade de pacotes, o que parece alinhado às constatações de Chen e Baker (2010).

Com base nessa variedade de estudos, é possível verificar que o tema dos pacotes lexicais tem mobilizado os linguistas aplicados. Não obstante haver muitas pesquisas no assunto, Byrd e Coxhead (2010) notam que acontece ainda pouca orientação

¹⁶⁸ O Br-ICLE é coordenado por Tony Berber Sardinha (BERBER SARDINHA, 2001).

acerca de abordagens pedagógicas eficazes para o ensino das combinações recorrentes de palavras (p.33). Em uníssono, Granger e Meunier (2008), por sua vez, apregoam que existe “necessidade urgente de mais evidência empírica do impacto real de uma abordagem fraseológica para o ensino e a aprendizagem”¹⁶⁹ (p. 249). Ao eleger os pacotes lexicais como forma de aproximação da linguagem, deseja-se unir esforços ao atendimento desse apelo.

Para concluir, vale ressaltar que o campo do Processamento de Línguas Naturais (PLN) vem empregando considerável esforço na empreitada da extração automática de *Multiword Expressions* (MWE) (a exemplo de RAMISCH *et al.*, 2013; VILLAVICENCIO *et al.*, 2010; CASELI *et al.*, 2009 e VILLAVICENCIO *et al.*, 2007). Especificamente, MWEs são expressões, cujas propriedades sintáticas ou semânticas não podem ser extraídas das partes (SAG. *et al.*, 2002, APUD VILLAVICENCIO *et al.*, 2007 p. 1035). Tais fenômenos de linguagem incluem uma variedade de ocorrências linguísticas, tais como *phrasal verbs* (verbos frasais, a exemplo de ‘*come up with*’), compostos nominais (‘*soccer ball*’) e expressões institucionalizadas (‘*rice and beans*’, ‘*bread and butter*’, ‘*ham and cheese sandwich*’), dentre outras. O recorte que define as margens desses fenômenos linguísticos, por sua vez, é distinto daquele dos pacotes lexicais em LC, independentemente de ambos serem passíveis de extração eletrônica. Não só aqueles admitem termos com duas palavras (‘*ad hoc*’ e ‘*be cool*’), como tendem a ser idiomáticos por força de sua definição (‘*rock the boat*’ e ‘*spill the beans*’). Portanto, embora estejamos tratando de agrupamentos recorrentes de palavras, as constituições das duas realizações de língua são distintas.

¹⁶⁹ “Urgent need for more empirical evidence of the actual impact of a phraseological approach to teaching and learning.” (GRANGER e MEUNIER, 2008, p. 249)

**II. SEGUNDA PARTE -
A APLICAÇÃO DO *CORPUS***

5. CAPÍTULO 5

“Onde não posso deixar de ser metódico é na dúvida.”

Casimiro de Brito

METODOLOGIA

Neste capítulo, apresenta-se a metodologia adotada para o estudo. Para iniciar, retomam-se os objetivos e as perguntas de pesquisa. Depois, descrevem-se o *corpus* adotado e os procedimentos organizacionais em relação ao acervo. Posteriormente, aborda-se a concepção teórico-analítica em termos da escolha da metodologia e da ferramenta eleita para a análise do *corpus*. Por fim, relata-se o passo a passo empírico para extração dos pacotes lexicais, momento em que também são feitas considerações acerca de decisões que deram formato ao trajeto investigativo.

5.1. Dos objetivos e das perguntas de pesquisa

Conforme apresentado no capítulo de introdução do trabalho, o tema desta investigação é o ensino da produção de *abstracts* no campo de conhecimentos da saúde em língua inglesa. O objetivo é descrever e analisar a constituição de um *corpus*, compilado a partir das áreas especializadas de Medicina, Nutrição e Farmácia. É igualmente meta do estudo, ao promover a sistematização didática das conclusões, contribuir com os avanços do campo de ensino e aprendizagem de EAP.

Vale rememorar que a adoção do gênero/registo discursivo consagrado para a tese deu-se em função da alta densidade da carga informacional em relação à pequena extensão da produção, aspectos que contribuem para a complexidade da redação e da própria compreensão do texto em questão. Como o *abstract* divulga a pesquisa científica e, portanto, promove, tanto o reconhecimento, quanto a credibilidade da ciência, duas premissas necessárias para o compartilhamento do conhecimento entre os membros da comunidade pesquisadora, mostra-se relevante investigá-lo. Para tanto, miram-se os modos de dizer dos autores dos textos, porque é importante entender como os cientistas do campo eleito para investigação selecionam e apresentam informações nos *abstracts* de seus trabalhos acadêmicos. Especificamente, o argumento é de que os padrões colocacionais representam um modelo implícito para as escolhas fraseológicas mais comuns, no sentido de que desvios no padrão habitual impactam os leitores e determinam mudanças nos paradigmas de pesquisa.

Para recortar o objeto de investigação, adota-se o posicionamento de que o texto especializado representa “o instrumento ou o resultado de uma atividade comunicativa

socioprodutiva especializada” (HOFFMANN, 1998, p. 77, APUD, KRIEGER & FINATTO, 2004, p. 113).

Para entender os padrões colocacionais do *corpus*, adotaram-se, como forma de aproximação, pacotes lexicais ou *lexical bundles* (BIBER *et al.*, 1999), grupos poli-lexicais identificados pela frequência de ocorrência, por serem desafiantes para o aprendiz, uma vez que funcionam como blocos de construção para a linguagem acadêmica. A extração de pacotes lexicais é um modo objetivo e direto de relacionar unidades multi-palavras com certo nível de estabilidade. Além disso, essa tarefa vai ao encontro da didática de ensino e aprendizagem de EAP, da ordem de descrever a linguagem adotada nos gêneros/registros característicos dos campos acadêmicos e especializados (BIBER e CONRAD, 2009, p. 3).

Para chegar às questões que se deseja responder com a investigação e que serão apresentadas no final desta subseção, retomam-se alguns pressupostos sobre escrita acadêmica e o funcionamento da língua que serviram de norte à concepção do trabalho:

1. A linguagem especializada funciona em blocos fixos (SINCLAIR, 1991) e se orienta por princípios maiores do que a livre escolha dos usuários;
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 escolha de conteúdo e estilo (SWALES, 1990, p. 58)”; ou seja, os objetivos comunicativos reconhecidos pelos especialistas influenciam o tema, o estilo 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). Isso porque 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, mas 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.

Por fim, a asserção fundamental desta pesquisa é de que aprender a combinar adequadamente as palavras para construir textos em um determinado gênero escrito ajuda o acadêmico a fazer escolhas linguísticas apropriadas e assim a desenvolver referências acerca do tipo de linguagem que se espera que produza.

Para dar conta do postulado estabelecido, 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. Outrossim, é importante ter acesso a trabalhos descritivos para verificação de aspectos que são constantes e uniformes, bem como para que se possa constatar pontos de variação, cotejando-os entre si, a fim de sistematizar as descobertas, para ensiná-las melhor. Para tanto, o estudo se organiza a partir das 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?

Destaque-se que, por palavras-tópico e elementos conexos, relacionam-se os termos centrais que compõem os pacotes lexicais e as demais palavras que a eles se associam na composição de cada estrutura.

Isso posto, é possível constatar que o estudo se organiza direcionado pelo *corpus*, já que não partimos de hipótese prévia acerca das principais associações lexicais, a serem verificadas na coleção de dados. Não obstante, vale ressaltar que a própria asserção contida na visão de linguagem, segundo a qual a língua varia, postulada na primeira questão, implica uma certa pressuposição acerca do *corpus*. De todo modo, é a partir da interpretação dos padrões gerados através do uso da ferramenta analítica que será dado conhecer a composição dos pacotes lexicais mais frequentes, para que seja possível saber exatamente o que se mostra prioridade categorizar e ensinar.

5.2. Do *corpus* do estudo

Foi objetivo do trabalho de concepção do *corpus* que o mesmo apresentasse representatividade, autenticidade, balanceamento e diversidade, aspectos que Aluísio e Almeida (2006, p. 158-159) destacam como relevantes para a organização de um acervo de dados. O objetivo desta seção é realçar o modo como a compilação da coleção de textos reflete essas características.

Para situar o primeiro critério de organização do *corpus*, é importante dizer que a concepção da pesquisa se originou da constatação de carência de suporte por parte de pesquisadores de uma universidade brasileira no campo das Ciências da Saúde, quanto à produção de *abstracts* em língua inglesa. Por esse motivo, a organização do acervo de dados contempla *abstracts* originários de diferentes gêneros/registros textuais. Tal diversidade visa a tornar o *corpus* representativo das produções do grupo prioritariamente beneficiário do estudo, bem como tem por objetivo dar conta das perguntas de pesquisa com mais robustez e promover uma sistematização didática mais ampla. Além disso, conforme mencionado anteriormente, Swales (1990) reconhece que um indivíduo não pertence a uma única comunidade discursiva. Daí a importância de desenvolver investigações que abranjam interfaces com múltiplos gêneros/registros textuais.

Igualmente, a coleção de dados deveria refletir, em alguma medida, a abrangência do campo da saúde. Determinou-se, por esse viés então, o recorte de incluir *abstracts* de sub-áreas do campo biomédico, sendo essas Medicina, Nutrição e Farmácia. Como o acervo foi especialmente compilado para o desenvolvimento da investigação, era preciso fazer escolhas em função do critério de representatividade aliado à acessibilidade e praticidade do *corpus*. Assim, as publicações internacionais adotadas foram selecionadas em função de os *abstracts* estarem disponíveis eletronicamente de forma gratuita, bem como, a partir do critério de avaliação *Qualis A*, pelos parâmetros da CAPES¹⁷⁰, relativo ao alto fator de impacto.

A disponibilidade de acesso também foi critério para a organização da parte nacional do acervo e remonta às dimensões de acessibilidade e praticidade aludidas há pouco. Está implícita aqui uma defesa à democratização da pesquisa acadêmica, de modo que seja

¹⁷⁰ Avaliação de 2014.

livremente passível de conhecimento e investigação, especialmente o *abstract*, cuja razão de ser, conforme amplamente discutido ao longo deste texto, é da ordem de promover o compartilhamento da investigação científica.¹⁷¹ Além do que, estando os dados disponíveis para uso público, não se torna necessário que o pesquisador solicite autorização para estudá-los.

Feitas essas considerações acerca dos critérios organizacionais, apresenta-se o *Corpus CISA* (nomeado pela aglutinação das palavras *Ciência e Saúde*), um acervo monolíngue especializado, compreendendo *abstracts* em língua inglesa das áreas de Medicina, Nutrição e Farmácia a partir de:

- a) Artigos acadêmicos publicados nos periódicos *Annals of Internal Medicine*¹⁷², *American Journal of Clinical Nutrition*¹⁷³ e *British Journal of Pharmacology*¹⁷⁴;
- 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;¹⁷⁵
- 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.

¹⁷¹ Contrariamente à expectativa, quando idealizado o *subcorpus* brasileiro, considerei, dentre as universidades públicas cujas produções integrariam o acervo, uma instituição de ensino superior que apresenta acervo intelectual (incluindo os *abstracts* e títulos das pesquisas) privado, ou seja, disponível apenas através de senha de acesso, de modo que foi necessário alterar o planejamento inicial.

¹⁷² <http://annals.org> (Último acesso em 20/12/2015.)

¹⁷³ <http://ajcn.nutrition.org> (Último acesso em 20/12/2015.)

¹⁷⁴ [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1476-5381](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1476-5381) (Último acesso em 20/12/2015.)

¹⁷⁵ 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.

Em relação à autoria, é possível reconhecer o *subcorpus* internacional como originário de usuários experientes. Isso porque o padrão de qualidade estabelecido para publicações especializadas de alto impacto determina que os textos apresentem um caráter léxico-gramatical de excelência, independentemente da relação dos usuários com a língua. Do mesmo modo, atribui-se o *subcorpus* brasileiro como originário de usuários de inglês como língua franca, pois não seria possível afirmar tratar-se de uma *corpus* de aprendizes, por exemplo, uma vez que os textos não se originam de um contexto de ensino/aprendizagem.

Por fim, apresentam-se os dados como integrantes de um *corpus* de estudo, uma vez que ambos os conjuntos de textos serão descritos neste capítulo, embora o *corpus* internacional seja de referência em relação ao nacional.

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

Tabela (4): *Corpus* CISA – *SubCorpus* Internacional

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

Fonte: Freitas (2016)

Tabela (5): *Corpus* CISA – *SubCorpus* Brasileiro

<i>Corpus</i> Brasileiro	<i>Abstracts</i>	Palavras	Tipos
Total	358	90072	9556
Artigos Publicados	118	19170	3852

Teses/Dissertações	140	46147	6299
Monografias	100	24755	4399

Fonte: Freitas (2016)

Para chegar ao montante de 180.170 palavras, que, de acordo com Aston (1997) e Berber Sardinha (2004), representa 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)¹⁷⁶. Do mesmo modo, foram analisados outros trabalhos significativos na exploração de *abstracts* e/ou de pacotes lexicais, com coleções de proporções semelhantes ou menores do que esta (a exemplo de BERBER SARDINHA, 2003; DAYRELL, 2009^a, DAYRELL e ALUÍSIO, 2008, FENG, 2006 e SÁNCHEZ HERNÁNDEZ, 2013).¹⁷⁷

Ademais, o gênero/registo discursivo adotado para este estudo contém um número reduzido de palavras. Especificamente, os *abstracts* do acervo compilado apresentam uma média de apenas 255 palavras por texto. No entanto, os traços linguísticos que se busca reconhecer, os pacotes lexicais, representam a sequência de palavras mais frequente de um registo (BIBER *et al.*, 2004). Portanto, esses aspectos atestam a favor da dimensão da coleção de dados estar em condições de dar conta das perguntas estabelecidas para a pesquisa.

O critério da quantificação dos acervos internacional e brasileiro, por seu turno, visou que fosse possível lidar com um número de palavras equilibrado em ambos os *corpora* de estudo. Para tanto, também a organização das coleções de dados foi balizada a partir de pesquisas que igualmente adotaram um critério de distribuição semelhante de palavras nos *subcorpora* cotejados, tais quais, Chen e Baker (2010), Dayrell e Aluísio (2008), Luzon Marco (2000) e Sánchez Hernández (2013). Além disso, buscou-se trabalhar com a mesma frequência bruta convertida para extração dos grupos multi-palavras em ambos os *subcorpora*, como será abordado ao longo do detalhamento dos procedimentos metodológicos.

Os *abstracts* selecionados para compor o acervo são do tipo informativo, ou seja, caracterizam-se pela natureza descritiva que visa à síntese do trabalho acadêmico

¹⁷⁶ “I certainly do not subscribe to the view that, *ceteris paribus*, the bigger a corpus the more representative it will be.” (SWALES, 2002, p. 223).

¹⁷⁷ Outrossim, findo este estudo, é intenção que o acervo passe a ser dinâmico, sofrendo acréscimos, para que seja possível tecer novas análises.

abordado. Ao mesmo tempo, uma peculiaridade de apresentação dos textos é que parte desses está em formato estruturado, representado por subtítulos introduzindo os segmentos de introdução, metodologia, resultados e discussão. Outra parte do acervo se apresenta em formato desestruturado, à moda de texto corrido.

Embora não seja objetivo desse trabalho cotejar o modelo CARS¹⁷⁸ (SWALES, 1990; SWALES e FEAK, 2009), que descreve os movimentos e passos argumentativos retóricos dos textos, cumpre dizer que parte dos *abstracts* do acervo nacional não contempla a descrição de todos os segmentos representativos da descrição de uma pesquisa acadêmica. Faz-se esse registro por questões etnográficas que se entendem pertinentes à descrição do *corpus* de uma investigação. Essas assimetrias, no entanto, são naturais de uma coleção de dados que representa uma porção de língua autêntica, portanto, variável, como bem observam Swales (2011) e Biber (2006), especialmente em se tratando do *subcorpus* que apresenta usuários em diferentes estágios de experiência em relação ao uso do gênero/registo em questão.

Em relação à captura do *corpus* propriamente, o acervo foi copiado texto a texto e salvo em formato *plain text* (.txt), havendo sido excluídos títulos, bem como as identificações de autoria. Cada *abstract* corresponde a um arquivo identificado pelo título original do trabalho, conforme colhido no site ou plataforma de origem. Arquivos em *word* detêm os títulos, numeração e períodos de inclusão para cada publicação, no caso do *corpus* internacional e para cada gênero de texto de cada universidade, no caso do acervo brasileiro, de acordo com o padrão a seguir:

Tabela (6): Padrão de Arquivamento da Identificação do SubCorpus Internacional

Nome da Revista	Ano de Publicação	Mês	Número do Texto	Título
-----------------	-------------------	-----	-----------------	--------

Fonte: Freitas (2016)

Tabela (7): Padrão de Arquivamento da Identificação do SubCorpus Brasileiro

Nome da Universidade	Gênero Original	Ano de Inclusão	Número do Texto	Título
----------------------	-----------------	-----------------	-----------------	--------

Fonte: Freitas (2016)

¹⁷⁸ *Create-A-Research-Space*.

Destarte, no caso dos periódicos internacionais, cada *subcorpus*, corresponde a uma pasta identificada pelo mês de publicação. Tais pastas foram salvas em uma pasta maior com o nome do periódico correspondente. Com relação à coleção brasileira, cada universidade corresponde a uma pasta, contendo subpastas relativas a cada gênero discursivo de publicação. Para fins de praticidade no manuseio dos *subcorpora*, foram organizadas novas pastas contendo, para o acervo internacional, três subpastas correspondentes a todos os textos de cada periódico (desta vez, sem divisões mensais ou arquivo em *word* contendo descrição dos dados). Para o acervo nacional, as pastas adicionais foram separadas por gênero de texto (igualmente, sem outros arquivos descritivos anexados).

5.3. Da concepção teórico-analítica

São dois os aspectos de que trataremos nesta parte: a escolha da linha metodológica, que afirma o nexos do percurso empírico, e a apresentação da ferramenta eleita para analisar os dados.

Uma vez que esta investigação tem por meta reconhecer padrões que norteiam o uso de recursos linguísticos adotados por grupos de pesquisadores em suas comunidades de prática, a partir da análise de dados autênticos e apoio do uso da tecnologia, elegeu-se a Linguística de *Corpus* (LC), para nortear a concepção metodológica do trabalho.

A adoção dos pressupostos da LC, face ao objetivo sintetizado, justifica-se com base em quatro características do campo descritas por Biber *et al.* (1998):

1. O componente empírico, ou seja, o experimentalismo na observação de padrões de língua em uso a serem analisados com base em contextos autênticos;
2. O trabalho com exemplares de língua representativos, arquivados em bases de dados eletrônicos;
3. A adoção de *software* para observar padrões linguísticos;
4. O uso de técnicas analíticas quantitativas e qualitativas para interpretar os resultados aferidos.

Em outras palavras, não se está falando em observar o que é possível teoricamente, mas em estudar o uso com base nos modos como a linguagem ocorre nos textos. O ponto de partida é a concepção da língua como um sistema de probabilidades, a partir do qual ressaltam-se as associações entre itens lexicais e seus contextos.

Ao reiterar-se adesão aos princípios da LC, é relevante destacar o papel das ferramentas computacionais como suporte essencial ao processo analítico. A esse respeito, o próprio Sinclair (2004) referiu que a essência do *corpus* em relação ao texto encontra-se no fato daquele não poder ser observado diretamente, já que a exploração dos dados é feita através de ferramentas, como os concordanciadores (p. 189).

Anthony (2013) reforça que “a funcionalidade oferecida pelas ferramentas de *software* determina amplamente quais métodos de pesquisa em LC estão disponíveis ao pesquisador”¹⁷⁹ (p. 141). A implicação é que as características da ferramenta analítica se mostram altamente importantes para o trabalho do linguista de *corpus*. Entretanto, Anthony (*op. cit.*) também ressalta o valor de se esclarecer a diferença entre *ferramentas de observação de dados de um corpus* e *o corpus em si*. Segundo ele, tal distinção tende, por vezes, a ter as fronteiras borradas no campo, porque as ferramentas analíticas são baseadas em *software* e, conseqüentemente, são abstratas. Elas dependem de linguagem de programação, de algoritmos, de métodos de armazenamento e codificação, assim como do desenho visual de interfaces. Tais aspectos escapam à maioria dos linguistas aplicados que temos foco na análise de fatores linguísticos *per se*. Não se pode omitir, tampouco, o fato de que pode haver divergências em análises de diferentes ferramentas. Isso porque elas podem ter sido alimentadas diversamente em relação a como examinar dados e, portanto, os *corpora* são processados diferentemente em termos, por exemplo, de uso de maiúsculas e minúsculas ou de formas contraídas de palavras.

Por fim, os *software* fornecem evidências que necessitam ser explicadas, daí Biber *et al.* (1998) acima referir a um cotejamento quantitativo e qualitativo para interpretação dos resultados. Todavia, cabe apontar que nem o trabalho computacional, nem o olhar do analista são neutros no processo analítico. Em detrimento da precisão da ferramenta adotada e da meticulosidade no trajeto analítico, as escolhas feitas ao longo do processo de dar conta das evidências, a própria decisão acerca do que analisar e do que deixar de fora, configuram posições interpretativas particulares.

¹⁷⁹ “The functionality offered by software tools largely dictates what corpus linguistics research methods are available to a researcher.”(ANTHONY, 2013, p. 141)

Elegeu-se o conjunto de ferramentas AntConc¹⁸⁰ (ANTHONY, 2014), desenvolvido por Lawrence Anthony, professor da Universidade de Waseda/Japão, primeiramente por se tratar de um *software* livre, o que democratiza a pesquisa em LC.

Além disso, são vantagens do programa que vem sendo adotado por linguistas de *corpus*, tradutores, professores de línguas adicionais e estudantes ao redor do mundo, os seguintes aspectos:

1. A pequena dimensão do arquivo (3,67MB) permite *download* rápido e ocupa pouco espaço em disco;
2. A execução é de simples processamento, bastando um duplo clique do *mouse*, ficando dispensados licenciamento e instalação;
3. O *freeware* apresenta versões para os sistemas *Windows*, *Mac OS X* e *Linux*;
4. A facilidade de uso permite acessar várias funcionalidades em uma mesma interface, com apenas um clique.

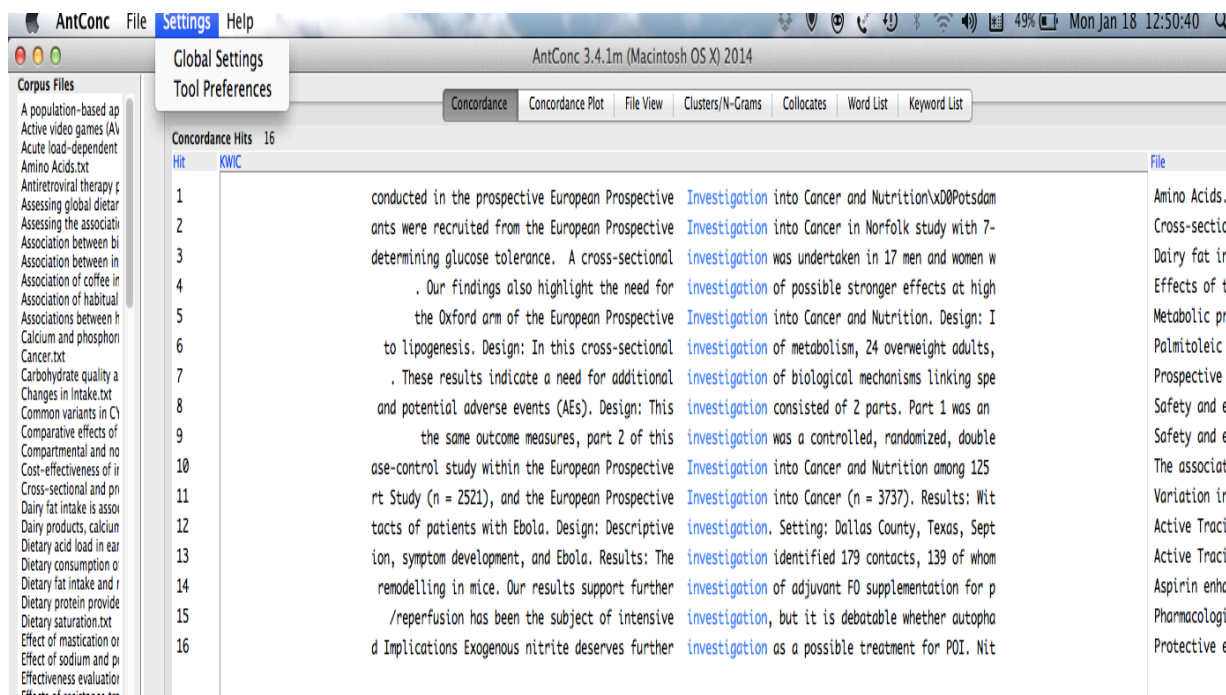
São igualmente características do AntConc permitir análise de textos de forma automática e desenvolver cálculos estatísticos. Esse último aspecto facilita imensamente a análise, uma vez que os fenômenos linguísticos encontrados através dos *corpora* apresentam dois aspectos vitais que o analista precisa levar em consideração, quais sejam, a conjugação dos quesitos frequência e dispersão. Se a ferramenta empregada já coteja esses aspectos, tem-se como ponto de partida uma medida representativa de precisão.

O sistema opera em língua inglesa ou em qualquer outra língua no padrão *Unicode*. Podem-se carregar coleções de textos simples ou anotados, e o software está habilitado para trabalhar com um número ilimitado de itens, desde que os arquivos sejam salvos em formato *plain text* (.txt).

A figura abaixo é a imagem da tela de abertura. Nela se pode visualizar a área de carregamento de arquivos à esquerda (*Corpus File*), uma área para configurações específicas (*Global Settings*), uma área de preferências da ferramenta (*Tool Preferences*), através da qual é possível inserir um *corpus* de referência, por exemplo, além das diferentes abas (de ‘*Concordance*’ à ‘*Keyword List*’), que correspondem às opções de ferramentas disponíveis.

¹⁸⁰ Disponível em <http://www.laurenceanthony.net/software.html> (Último acesso em 28/12/2015).

Figura (2): AntConc, Tela de Abertura



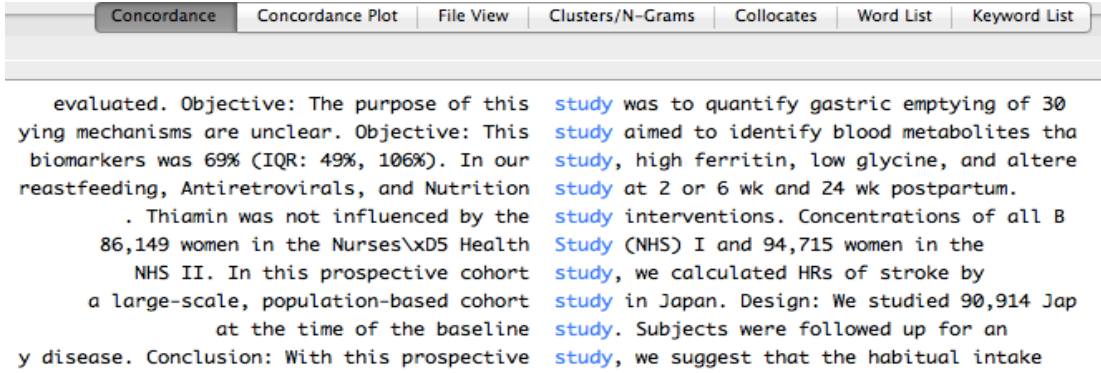
Fonte: Freitas (2016)

Para iniciar o trabalho, carregam-se os textos do *corpus* a partir do menu *File* (*Open Dir*, para carregar um diretório). Ao clicar-se, abre-se uma janela de navegação e, uma vez selecionada a pasta que se deseja carregar, basta clicar em ‘ok’ e os textos são automaticamente carregados. Os arquivos são visualizados no quadro *Corpus Files*. Sempre que quisermos salvar os resultados, clica-se em *File* e seleciona-se *Save Output to Text File*, e os resultados serão salvos em formato *bloco de notas*.

O lado direito da tela é onde os resultados aparecem. Seguindo da esquerda para a direita, a partir da tela de abertura (Figura (2) acima), temos disponíveis as seguintes ferramentas:

1. *Concordance* – possibilita mostrar um termo ou sequência de palavras e as respectivas linhas de concordância;

Figura (3): Concordance Tool

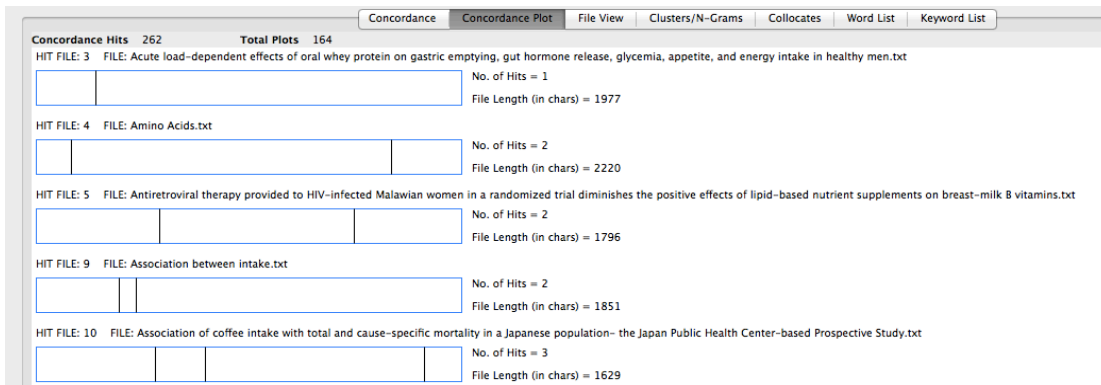


Fonte:

Freitas (2016)

2. *Concordance Plot* – oportuniza exibir os resultados do termo ou sequência de palavras em ‘concordance’, em códigos de barras que apontam em que arquivo(s) do *corpus* encontram-se as ocorrências e sua localização espacial. Ao clicar-se nas linhas dos códigos de barra, visualizam-se os respectivos arquivos;

Figura (4): *Concordance Plot Tool*

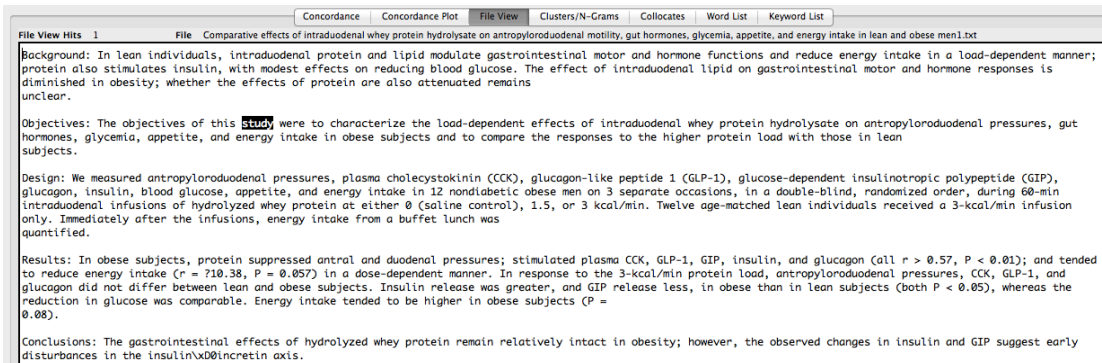


Fonte:

Freitas (2016)

3. *File View* – propicia a visualização individual dos textos que formam o *corpus*;

Figura (5): *File View Tool*



Fonte:

Freitas (2016)

5. *Clusters/N-Grams* - Faz buscas de *clusters/n-grams*, permitindo a listagem ou do termo ou de associação de palavras combinando-o(s) a 2 ou mais palavras que ocorrem em determinada frequência ou – o caso de nosso interesse maior – permite resgatar pacotes lexicais (*lexical bundles*), combinações de palavras recorrentes em um *corpus*, especificando tamanho mínimo e máximo (número de palavras) por n-grama, frequência mínima e dispersão,¹⁸¹

Figura (6): *Clusters/N-Grams Tool*

Total No. of N-Gram Types 34				Total No. of N-Gram Tokens 147	
Rank	Freq	Range	N-gram		
1	12	12	of this study was		
2	12	12	this study was to		
3	8	8	was to evaluate the		
4	6	5	rio grande do sul		
5	6	6	study was to evaluate		
6	6	6	the aim of this		
7	6	6	the objective of this		
8	5	5	aim of this study		
9	4	4	evaluate the effect of		
10	4	4	it is possible to		
11	4	4	objective of this study		

Fonte: Freitas (2016)

5. *Collocates* – proporciona acesso aos colocados de um termo e permite investigar padrões não sequenciais;

Figura (7): *Collocates Tool* - Exemplo a partir do termo de busca ‘research’

¹⁸¹ Na próxima seção, serão detalhadas as escolhas de extensão, frequência e dispersão estabelecidas para chegar aos pacotes lexicais do *Corpus CISA*.

Total No. of Collocate Types: 99		Total No. of Collocate Tokens: 159			
Rank	Freq	Freq(L)	Freq(R)	Stat	Collocate
1	1	0	1	10.22656	triangle
2	1	1	0	10.22656	stand
3	1	0	1	10.22656	scholarships
4	1	0	1	10.22656	proper
5	1	0	1	10.22656	projects
6	1	0	1	10.22656	park
7	1	0	1	10.22656	overseas
8	1	1	0	10.22656	opportunities
9	1	0	1	10.22656	north
10	1	0	1	10.22656	investigators
11	1	1	0	10.22656	glaxosmithkline
12	1	1	0	10.22656	framework

Fonte: Freitas (2016)

6. *Word List* – possibilita a geração de uma lista de todas as palavras (*tokens*) e de todos os tipos (*types*)¹⁸² do *corpus*, apresentadas em ordem de frequência, em ordem alfabética ou a partir das terminações;

Figura (8): *Word List Tool*

Word Types: 3852		Word Tokens: 19170		Search Hits: 0	
Rank	Freq	Word	Lemma Word Form(s)		
1	1245	the			
2	899	of			
3	790	and			
4	538	in			
5	384	to			
6	322	a			
7	257	was			
8	212	with			
9	184	were			
10	183	for			
11	137	is			
12	132	as			

Fonte:

Freitas (2016)

6. *Key Word List* – faculta comparações entre um *corpus* de estudo contrastado a um *corpus* de referência, gerando uma lista de palavras-chave.

Figura (9): *Key Word List Tool*

¹⁸² No exemplo, “*I came, I saw, I concordanced*”, tem-se 6 palavras, que representam 6 *tokens*, portanto. Mas 3 delas são repetidas, portanto, há 4 *types*.

Concordance		Concordance Plot		File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Types Before Cut: 41491		Types After Cut: 29571		Search Hits: 0				
Rank	Freq	Keyness	Keyword					
34	212	301.976	stood					
35	284	301.939	feet					
36	223	299.136	surface					
37	877	295.005	each					
38	1445	293.863	did					
39	1775	288.995	could					
40	321	276.209	door					
41	788	275.906	men					
42	622	275.289	came					

Fonte: Freitas (2016)

Para realizar uma nova pesquisa, não é necessário fechar o aplicativo, basta limpar as informações anteriores em *File* (Figura (2)) que abrirá várias opções, dentre elas *Clear All Tools and Files*.

Apresentada uma síntese dos recursos da ferramenta analítica,¹⁸³ passamos às etapas da análise.

5.4. Das etapas da análise

Para abrir esta seção, chama-se a atenção do leitor para o modo como escolheu-se encaminhar a narrativa, já que, para analisar o *corpus*, foi necessário, naturalmente, tomar decisões que deram formato específico ao trajeto investigativo. Optou-se por lidar com esses pontos nesta altura do texto, em vez de fazê-lo nas seções anteriores, para oferecer maior riqueza à leitura do que se nos atívéssemos, estritamente, à sequência do passo a passo analítico. Por isso, apresenta-se a seção em três partes:

- (1.) a definição de critérios para identificação dos pacotes lexicais e para o tratamento analítico das ocorrências;
- (2.) o roteiro de cotejamento dos *subcorpora* e
- (3.) o passo a passo analítico.

¹⁸³ Para uma apresentação mais detalhada do conjunto de ferramentas, há tutoriais apresentados pelo autor em <https://www.youtube.com/watch?v=9TsqFVrUY00> (Último acesso em 03/04/18).

5.4.1. Dos critérios de identificação e tratamento dos pacotes lexicais

Para o trabalho de extração dos pacotes lexicais, foi necessário, primeiramente, estabelecer critérios para a extensão e extração das ocorrências, uma vez que os grupos poli-lexicais adotados como unidade de análise são sequências pré-fabricadas, baseadas em número de palavras associadas, frequência de ocorrência no *corpus* e princípios distributivos.

Para o estabelecer a extensão, adotou-se o critério de quatro palavras gráficas, a exemplo de Biber *et al.* (2004), Biber e Barbieri (2007), Dayrell (2009^a), Chen e Baker (2010), Cortes (2004) e Hyland (2008^a, 2008^b). Seguiu-se o princípio norteador de que os pacotes lexicais de três palavras são muito frequentes e geralmente se abrigam em sequências de quatro palavras¹⁸⁴. Já pacotes menores do que os de três palavras podem ser fragmentos de texto apenas, enquanto os mais longos tendem a ser mais escassos.

Para determinar a frequência, ponderou-se sobre a extensão do *corpus*, tendo em vista que o ponto de corte de afluência é estabelecido de modo arbitrário (BIBER *et al.*, 2004, p. 376). Biber *et al.* (1999), por exemplo, especifica que os pacotes lexicais estariam restritos a combinações recorrentes verificáveis pelo menos 10 vezes por milhão de palavras, na proporção de dispersão de 5 textos diferentes. Outros autores como Cortes (2004) e Hyland (2008^a, 2008^b e 2008^c) trabalharam com 20 ocorrências por milhão de palavras, cotejadas a uma dispersão de 10% dos textos. Já Biber e Barbieri (2007), estabeleceram o limite de 40 ocorrências por milhão de palavras. Para adequar a busca ao acervo, estabeleceu-se o ponto de corte de 5 ocorrências para 180.170 palavras, também 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). Neste caso, adotou-se um fator mais alto do que o preconizado por Biber *et al.* (*op. cit.*), o que, por sua vez, confere ainda maior segurança quanto à medida de força das associações de palavras a serem resgatadas no *corpus*, em relação à probabilidade de ocorrência.

Para o fator de dispersão, tomaram-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* (internacional e brasileiro).

¹⁸⁴ A exemplo de ‘*as a consequence of*’ que abriga ‘*as a consequence*’.

Por fim, para o trabalho de análise dos pacotes lexicais, decidiu-se por estudar as ocorrências nos eixos léxico-gramatical e funcional, como é tradição nas investigações do tema. Para fins do primeiro eixo, partiu-se das categorias postuladas por Biber *et al.* (1999), apresentadas no capítulo anterior. O trabalho primou por entender as ocorrências, seus encaixes e distinções inicialmente. Só depois disso, resgataram-se as classificações. Alguns novos padrões naturalmente foram percebidos e descritos, como será possível constatar no próximo capítulo.¹⁸⁵

Para a distribuição funcional, entretanto, optou-se por estabelecer uma nova classificação, uma vez que a categorização de Biber *et al.* (2004), que igualmente foi descrita no capítulo anterior, é complexa e tem por objetivo contemplar o trabalho do analista exclusivamente. O propósito aqui foi de estabelecer uma rota própria, com base no que dizem os dados, a fim de adotar o eixo funcional – assim como o léxico-gramatical – como possibilidade de suporte ao ensino posteriormente. Em outras palavras, a ideia é de que o aprendiz seja levado a perceber como os pacotes lexicais trabalham em termos estruturais e pragmáticos para dar sentido aos textos acadêmicos.¹⁸⁶

5.4.2. Do cotejamento dos *subcorpora*

A investigação do acervo visa a promover 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 explora os dados com vistas a analisar as diferenças de frequência dos padrões lexicais entre as partes componentes do *corpus*, especialmente, para contribuir para a qualificação da produção acadêmica de *abstracts* de teses, dissertações e monografias. Para tanto, adotou-se a seguinte sequência analítica:

1. Análise do *subcorpus* internacional;
2. Análise do *subcorpus* nacional como um todo;
3. Análise das publicações do *subcorpus* nacional;

¹⁸⁵ Agradeço ao Prof. Dr. Douglas Biber e à Profa. Dra. Randi Reppen, da *Northern Arizona University/USA*, pelo interesse e valiosa oportunidade de discussão em Lancaster/UK, no mês de julho de 2015, a respeito do modo de categorizar os pacotes lexicais deste estudo e pelo encorajamento a *encontrars meus próprios padrões nos dados*.

¹⁸⁶ Uma vez que a categorização funcional se organiza, especificamente, com base nos dados, a mesma será relacionada no próximo capítulo.

4. Análise das teses e dissertações do *subcorpus* nacional;
5. Análise das monografias de conclusão de curso do *subcorpus* nacional;
6. Análise das partes do *subcorpus* nacional em relação ao *subcorpus* nacional como um todo;
7. Análise comparativa do *subcorpus* internacional e nacional.

5.4.3. Do passo a passo analítico

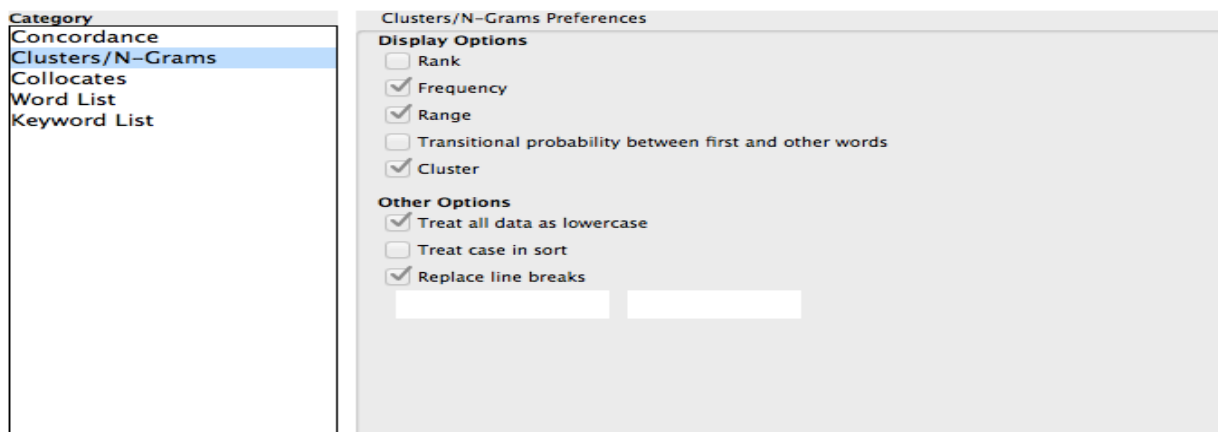
Cada passo do trabalho analítico compreendeu um número de etapas descritas abaixo:

Passo 1: O primeiro passo da análise foi extrair a lista dos pacotes lexicais mais frequentes e pedagogicamente significativos no *subcorpus* internacional e classificá-los. Esses pacotes são referidos por Cortes (2004), como *pacotes-alvo*¹⁸⁷. A tarefa compreendeu as seguintes etapas:

1ª Etapa: Após carregar o *subcorpus* internacional no AntConc, clicou-se na opção *Cluster/N-Gram* na caixa de ferramentas, escolhendo a opção ‘*Treat all data as lower case*’ em ‘*Tool Preferences*’, a fim de transformar todas as palavras em minúsculas. O objetivo foi o de não incorrer no risco de uma sequência de palavras contar duplamente, dependendo de sua localização na frase. Em ‘*Tool Preferences*’, também optou-se por ranquear os grupos de palavras a serem gerados pelos critérios ‘*rank*’, ‘*frequency*’ e ‘*range*’. Por fim, na tela principal, marcaram-se as opções de extensão mínima e máxima em ‘4’, a frequência mínima em ‘5’ e a dispersão em ‘5’;

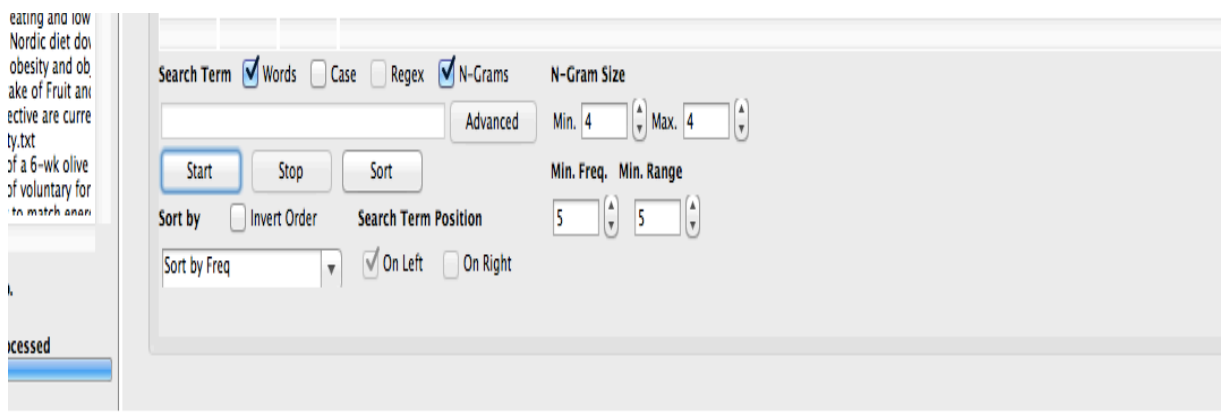
Figura (10): AntConc, Caixa de Ferramentas

¹⁸⁷ *Target bundles*.



Fonte: Freitas (2016)

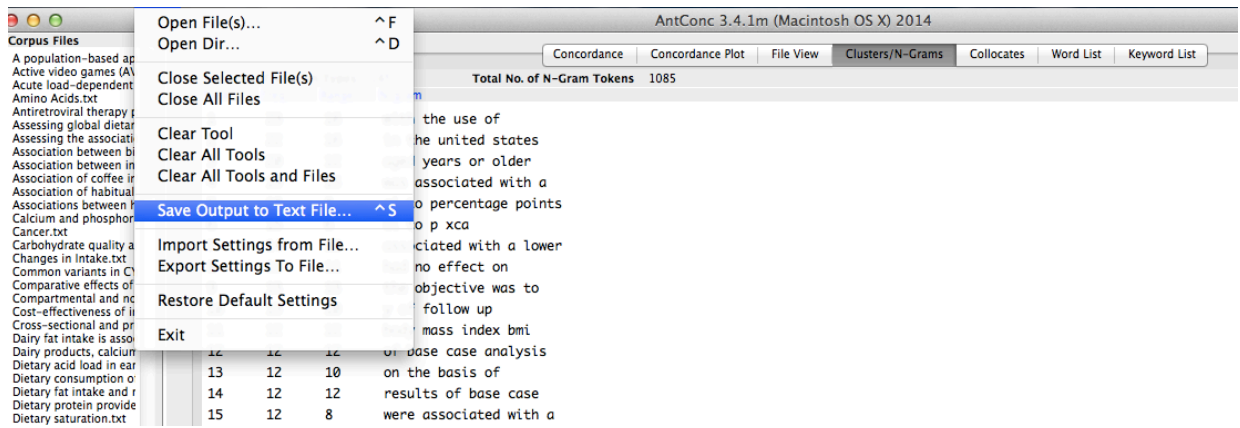
Figura (11): AntConc, Extensão, Frequência e Dispersão



Fonte: Freitas (2016)

2ª Etapa: Gerada a lista dos pacotes lexicais, os mesmos foram destacados e salvos pelo uso do comando ‘File’ e ‘Save Output to Text File’. Esse procedimento facilitou a seleção manual da relevância pedagógica das colocações estendidas. Sequências de palavras que incluíam termos técnicos encriptados, a exemplo de ‘*ci to p xca*’ ou sequências de contraste, como ‘*energy x ray absorptiometry*’ foram excluídas;

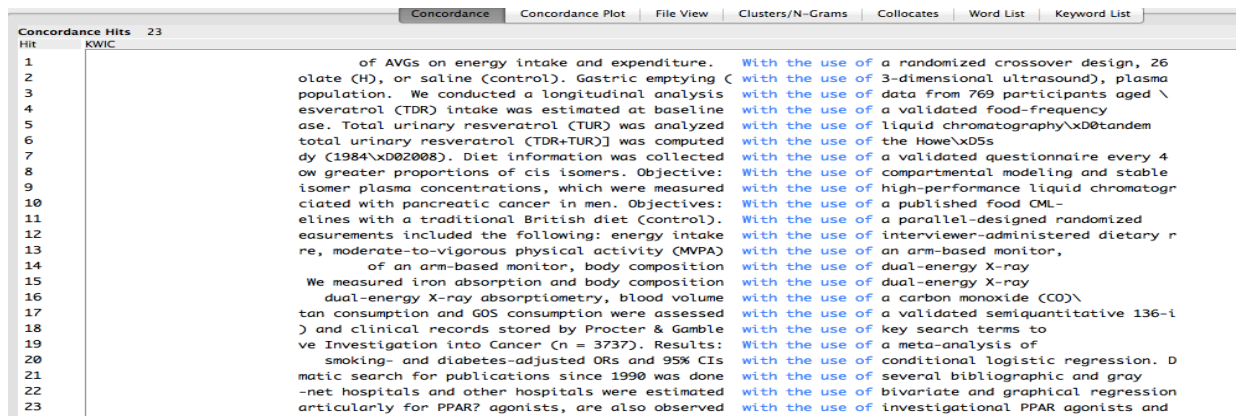
Figura (12): AntConc, Função de Salvamento de Resultados



Fonte: Freitas (2016)

3ª Etapa: Extraídas e selecionadas as ocorrências, essas foram analisadas e classificadas quanto às características léxico-gramaticais e funcionais. Para tanto, voltou-se à ferramenta e visualizou-se uma a uma, clicando na ocorrência, a fim de cotejá-la à linha de concordância. Essa tarefa facilitou, especialmente, o trabalho de classificação funcional. Essa etapa também demandou o estabelecimento de categorias. Para essa parte do trabalho, imprimiu-se a lista de ocorrências, que foram recortadas e criaram-se conjuntos. Tendo em vista a meta de estabelecer categorias, foi um exercício produtivo.

Figura(13) AntConc – Pacote Lexical Relacionado à Linha de Concordância



Fonte: Freitas (2016)

Passo 2: O passo seguinte foi a extração, seleção e análise dos pacotes lexicais da porção brasileira do *corpus*. Para dar conta da tarefa, repetiram-se as etapas 1, 2 e 3 descritas. Vale dizer que, por questões de facilidade de manuseio e preferência, optou-se por trabalhar com o AntConc em duas janelas, ou seja, o *software* foi baixado duas vezes e operou-se, simultaneamente, com duas coleções de dados sem precisar limpar a ferramenta.

Embora não seja necessário ao analista adotar essa metodologia, pois o AntConc permite clonar os resultados em uma janela menor, a mesma se provou assaz prática.

Passo 3: Em segmento, re-fracionou-se o acervo brasileiro, que novamente é reproduzido abaixo, com base no tipo de produção.

Tabela (5): Corpus CISA – SubCorpus Brasileiro

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

Fonte: Freitas (2016)

O trabalho se processou na seguinte sequência:

Etapa 1: Primeiramente, carregaram-se, em três rodadas, cada uma as pastas do *subcorpus* brasileiro, seguindo as etapas de ajuste da ferramenta *Clusters/N-Gram* conforme descrição acima e, posterior identificação dos pacotes lexicais;

Figura (14): AntConc - Extração dos Pacotes Lexicais do SubCorpus Brasileiro de Publicações

Total No. of N-Gram Types 8				Total No. of N-Gram Tokens 61
Rank	Freq	Range	N-gram	
1	12	12	of this study was	
2	12	12	this study was to	
3	8	8	was to evaluate the	
4	6	5	rio grande do sul	
5	6	6	study was to evaluate	
6	6	6	the aim of this	
7	6	6	the objective of this	
8	5	5	aim of this study	

Fonte: Freitas (2016)

Figura (15): AntConc - Extração dos Pacotes Lexicais do SubCorpus Brasileiro de Teses/Dissertações

Total No. of N-Gram Types				32	Total No. of N-Gram Tokens				272
Rank	Freq	Range	N-gram						
1	18	18	of this study was						
2	18	18	this study was to						
3	17	17	the aim of this						
4	14	14	aim of this study						
5	14	14	as well as the						
6	12	12	complete abstract click electronic						
7	11	11	abstract click electronic access						
8	11	11	click electronic access below						
9	11	8	in relation to the						
10	9	9	of the present study						
11	9	8	on the other hand						
12	8	7	in the presence of						
13	8	8	the present study was						
14	8	8	this study aimed to						
15	8	8	was to evaluate the						
16	7	6	for the treatment of						
17	7	5	it was observed that						
18	7	7	study was to evaluate						
19	7	7	these results suggest that						
20	6	5	for the development of						
21	6	6	present study was to						
22	6	6	the analysis of the						
23	5	5	cl x nicas de						
24	5	5	in this study the						
25	5	5	is one of the						
26	5	5	nicas de porto alegre						
27	5	5	objective of this study						
28	5	5	one of the most						
29	5	5	study was to investigate						
30	5	5	the objective of this						
31	5	5	the other hand the						

Fonte: Freitas (2016)

Figura (16): AntConc - Extração dos Pacotes Lexicais do SubCorpus Brasileiro de Monografias

Total No. of N-Gram Types				10	Total No. of N-Gram Tokens				64
Rank	Freq	Range	N-gram						
1	10	6	in of the cases						
2	7	6	of this study was						
3	7	5	the aim of this						
4	6	5	as well as the						
5	6	6	hospital infantil joana de						
6	6	6	joana de gusma o						
7	6	6	the analysis of the						
8	6	6	this study was to						
9	5	5	in relation to the						
10	5	5	the objective of this						

Fonte: Freitas (2016)

Etapa 2: Em seguida, planilharam-se os resultados, relacionando-os aos achados gerais do *subcorpus* maior.

Etapa 3: Por fim, teceram-se análises comparativas para cada uma das pastas do *subcorpus* brasileiro ao *subcorpus* nacional como um todo e ao *subcorpus* internacional.

Passo 4: Em sequência, cotejaram-se os dois *subcorpora*, a fim de observar semelhanças e diferenças. Note-se que, mesmo comparando *corpora* diferentes neste momento, para a realização deste passo, não foi necessário normalizar a frequência através de cálculo estatístico, porque foram adotados *corpora* do mesmo tamanho. A tarefa se desenvolveu em duas etapas:

1ª Etapa: A primeira etapa do trabalho consistiu em análise quantitativa e das características dos pacotes lexicais dos dois conjuntos de dados;

2ª Etapa: A etapa seguinte foi a vez do estudo comparativo em termos das semelhanças e diferenças estruturais e funcionais entre os dois acervos.

Com base nos passos e etapas descritos, passamos ao capítulo seguinte em que apresentam-se as ocorrências estudadas.

6. CAPÍTULO 6

“All research – whether corpus based of any other kind requires interpretation. The presentation of evidence, application of theoretical principles, and development of an argument make interpretation in research studies different from intuition.”

Susan Conrad (2011, p. 52)

‘THESE DATA SUGGEST THAT...’: O QUE REVELAM OS DADOS

Neste capítulo, introduzem-se, primeiramente, as ocorrências da coleção internacional e, em sequência, os dados da coleção nacional como um todo. Inicia-se a exposição com os principais pacotes lexicais de cada acervo e, em segmento, introduzem-se as respectivas classificações de caráter funcional e léxico-gramatical. Para apresentar a distribuição funcional dos dados, preliminarmente, descreve-se a taxonomia desenvolvida para dar conta dessa tarefa. Posteriormente, a fim de destacar os padrões que tipificam o *subcorpus* brasileiro, expõe-se a constituição das associações recorrentes de palavras que integram as coleções de *abstracts* que compõem essa parte do *corpus*.

As análises tecidas nesta parte do texto serão retomadas de modo reflexivo no capítulo seguinte, quando do trabalho com a síntese analítica e as questões de pesquisa.

6.1. Do *subcorpus* internacional

Para a extração dos grupos poli-lexicais, conforme especificado no capítulo anterior, foi adotado o critério de extensão de sequências de 4 palavras gráficas. Para os fatores de frequência e dispersão, tomou-se o indicador ‘5’ como linha de corte. Com base nessa métrica, após o trabalho de refinar as ocorrências, chegou-se ao montante de 96 pacotes lexicais em 672 realizações individuais, em um acervo de 90.098 palavras. A ocorrência mais frequente foi ‘*with the use of*’ com 23 realizações em 16 textos. As 10 colocações estendidas de maior emprego ocorrem 142 vezes em 119 textos e se encontram relacionadas na Tabela 8 abaixo.

Os feixes lexicais relacionados a esta porção do *corpus* confirmam o princípio segundo o qual pacotes lexicais de 3 palavras são muito frequentes e geralmente se abrigam em sequências de 4 palavras (BIBER e BARIBIERE, 2007; BIBER *et al.*, 1999; CORTES, 2004). A fim de ilustrar esse princípio, na Tabela 8 abaixo, relacionam-se as primeiras ocorrências de cada *subcorpus*. Vale dizer que os cortes estabelecidos (+) representam possibilidades de associações para as sequências de palavras, com relação às palavras-tópico, termos centrais das colocações estendidas, e seus respectivos elementos conexos.

Tabela (8): Exemplificando Pacotes Lexicais de 3 Palavras Abrigados em Segmentos de 4 Palavras

Sequência	Frequência	Distribuição	Ocorrência
1	23	16	<i>with + the use of</i>
2	19	15	<i>was associated with + a</i>
3	14	10	<i>associated with a + lower</i>
4	13	11	<i>had + no affect on</i>
5	13	13	<i>the + objective was to</i>
6	12	12	<i>body mass index + bmi</i>
7	12	12	<i>of + base case analysis</i>
8	12	10	<i>on + the basis of</i>
9	12	12	<i>results + of base case</i>
10	12	8	<i>were associated with + a</i>

Fonte: Freitas (2016)

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. Foram consideradas as duas instâncias analíticas eleitas para tanto: a) as funções retóricas e b) as peculiaridades léxico-gramaticais.

Diferentemente da gama de trabalhos que se voltam para o tema e que, da mesma forma, adotam pacotes lexicais como unidade de aproximação do *corpus*¹⁸⁸, inicia-se a apresentação da distribuição dos agrupamentos de palavras pela classificação pragmática¹⁸⁹. A razão para esse posicionamento é o entendimento dessa caracterização como mais efetiva para a apreensão da contribuição das sequências multi-palavras para o todo dos textos. Afinal, é a partir do aporte de significado contextual que se torna possível extrair o sentido mais amplo das ocorrências nos *abstracts* e, com isso, captar o modo como o componente fragmentário organiza o todo.

Com base nessa lógica, para chegar à classificação funcional, foram estabelecidas categorias, especificamente, a partir da análise das linhas de concordância dos feixes de palavras e dos respectivos arquivos de texto. O critério para chegar aos 7 tipos

¹⁸⁸ Relacionados ao longo deste texto no cap. 4, seção 4.3.

¹⁸⁹ Vale lembrar que, por classificação pragmática ou funcional, entende-se o significado e os propósitos da linguagem. Conforme mencionado no cap. 4, subseção 4.4.4, tais aspectos visam a conferir textura ao discurso ou a organizar o texto de acordo com as situações e o contexto de uso.

estabelecidos foi da ordem de adotar uma divisão que fosse passível tanto para o trabalho de pesquisa, quanto para apoiar o ensino, conforme será abordado posteriormente. Toda categorização, bem sabemos, é sempre uma redução que implica escolhas e renúncias.

Chegou-se à taxonomia apresentada a seguir pelo sentido estabelecido a partir das associações lexicais, com base nas palavras-tópico e, por vezes, pelo sentido expresso pela representação de uma ou mais palavras do pacote especificamente. Os exemplos que se seguem às categorias, por sua vez, partem do *Corpus CISA*:

1. Estabelecer Comparação / Associação / Possibilidade / Tecer Avaliação / Quantificar – para agrupar ocorrências que estabelecem algum tipo de mensuração (ex.: *compared with the lowest, was associated with a, it is possible to, an important role in, little is known about, etc.*);
2. Apontar Resultados / Conclusões / Evidências / Recomendações / Consequências / Indícios dos Dados – para classificar os agrupamentos de caráter conclusivo (ex.: *these results suggest that, we concluded that, it was observed that, it's necessary to, etc.*);
3. Expressar Objetivos / Propósito – para reunir as sequências recorrentes com sentido em metas de um estudo (ex.: *the objective was to, our goal was to, etc.*);
4. Topicalizar – para fazer referência ao elemento temático (ex.: *prospective investigation into cancer, mass index in kg, etc.*);
5. Esclarecer Aspectos Metodológicos – para relacionar aspectos acerca de metodologia (ex.: *randomized controlled trials rct, hazards models were used, etc.*);
6. Organizar o Relato de Pesquisa – organizar as informações acerca do estudo (ex.: *in the absence of, in the presence of, etc.*);
7. Contextualizar / Situar – para agrupar colocações estendidas que situam o leitor em relação ao estudo (ex.: *trial was registered at, the cost effectiveness of, etc.*).

Isso posto, apresentam-se os grupos poli-lexicais desta porção do acervo organizadas com base em suas funções retóricas:

Tabela (9): Ocorrências do Subcorpus Internacional por Classificação Funcional

Categoria Funcional	Frequência	Distribuição	Ocorrência
---------------------	------------	--------------	------------

Estabelecer Comparação/ Associação /Possibilidade/ Tecer Avaliação /Quantificar	19	15	<i>was associated with a</i>
	14	10	<i>associated with a lower</i>
	13	11	<i>had no effect on</i>
	12	8	<i>were associated with a</i>
	9	7	<i>was associated with higher</i>
	9	6	<i>with a lower risk</i>
	8	6	<i>a lower risk of</i>
	8	6	<i>was associated with lower</i>
	7	7	<i>assessed risk of bias</i>
	7	5	<i>intake was associated with</i>
	7	5	<i>associated with increased risk</i>
	7	7	<i>was not associated with</i>
	6	5	<i>compared with the lowest</i>
	6	6	<i>was inversely associated with</i>
	6	6	<i>is associated with a</i>

	6	6	<i>a wide range of</i>
	5	5	<i>has been associated with</i>
	5	5	<i>have been associated with</i>
	5	5	<i>on the association between</i>
	8	7	<i>benefits and harms of</i>
	6	6	<i>an important role in</i>
	6	6	<i>little is known about</i>
	5	5	<i>plays an important role</i>
	5	5	<i>were estimated by using</i>
	5	5	<i>was assessed by using</i>
Apontar Resultados/ Conclusões /Evidências/ Recomendações/ Consequências/Indícios dos Dados	12	12	<i>results of base case</i>
	11	11	<i>results of sensitivity analysis</i>
	7	7	<i>these results suggest that</i>
	6	6	<i>the primary outcome was</i>
	5	5	<i>these data suggest that</i>
Expressar Objetivo /Propósito	13	13	<i>the objective was to</i>
	11	10	<i>for the</i>

			<i>treatment of</i>
	9	9	<i>this study was to</i>
	9	9	<i>of this study was</i>
	8	7	<i>in this study we</i>
	7	7	<i>studies are needed to</i>
	7	7	<i>the aim was to</i>
	6	6	<i>regression was used to</i>
	6	6	<i>was to examine the</i>
	6	6	<i>investigated the effects of</i>
	5	5	<i>aim of this study</i>
	5	5	<i>in this review we</i>
	5	5	<i>estimate the cost effectiveness</i>
	5	5	<i>examine the association between</i>
	5	5	<i>models were used to</i>
	5	5	<i>the aim of this</i>
	5	5	<i>tested the hypothesis that</i>
	5	5	<i>to estimate the cost</i>
	5	5	<i>to examine the</i>

			<i>association</i>
	5	5	<i>used to evaluate the</i>
	5	5	<i>the present study was</i>
	5	5	<i>was to investigate the</i>
Topicalizar	12	12	<i>body mass index bmi</i>
	8	8	<i>preventive services task force</i>
	7	7	<i>assessment of insulin resistance</i>
	7	7	<i>and body mass index</i>
	7	7	<i>model assessment of insulin</i>
	7	6	<i>quality adjusted life years</i>
	6	6	<i>body mass index in</i>
	6	6	<i>homeostasis model assessment of</i>
	6	6	<i>hepatitis c virus hcv</i>
	6	6	<i>mass index in kg</i>
	6	6	<i>reactive oxygen species ros</i>
	5	5	<i>and reference lists study</i>
	5	5	<i>c virus hcv</i>

			<i>infection</i>
	5	5	<i>prospective investigation into cancer</i>
Esclarecer Aspectos Metodológicos	12	12	<i>of base case analysis</i>
	12	11	<i>were randomly assigned to</i>
	8	8	<i>register of controlled trials</i>
	7	7	<i>incremental cost effectiveness ratios</i>
	7	7	<i>randomized controlled trials rcts</i>
	7	7	<i>years of follow up</i>
	6	6	<i>in vitro and in</i>
	6	6	<i>selection randomized controlled trials</i>
	6	6	<i>study selection randomized controlled</i>
	6	6	<i>cox proportional hazards models</i>
	5	5	<i>hazards models were used</i>
	5	5	<i>case base analysis the</i>
	5	5	<i>proportional</i>

			<i>hazards models were</i>
	5	5	<i>vitro and in vivo</i>
	5	5	<i>of bias data synthesis</i>
	5	5	<i>person years of follow</i>
Organizar o Relato de Pesquisa	23	16	<i>with the use of</i>
	12	10	<i>on the basis of</i>
	8	8	<i>the use of a</i>
	7	5	<i>in the presence of</i>
	5	5	<i>in the absence of</i>
Contextualizar/Situar	11	10	<i>the american college of</i>
	10	6	<i>the cost effectiveness of</i>
	8	8	<i>cochrane central register of</i>
	8	8	<i>central register of controlled</i>
	8	8	<i>American college of physician</i>
	6	6	<i>trial was registered at</i>
	5	5	<i>European prospective investigation into</i>
	5	5	<i>and strength of evidence</i>

	5	5	<i>this trial was registered</i>
--	---	---	----------------------------------

Fonte: Freitas (2016)

Saliente-se que, ao longo do trabalho de classificação, alguns pacotes apresentaram possibilidade de associação a mais de um grupo dentre as opções estabelecidas. A propósito dessa questão, Beber Sardinha e Dutra (20015) também constataram a mesma problemática em relação a um *corpus* de produções acadêmicas de aprendizes de língua inglesa. Nesses casos, busquei analisar as ocorrências tendo por base a abertura das linhas de concordância, valendo esse critério para ambos os *subcorpora*. Segui esse princípio a fim de desambiguar, especialmente, ocorrências das categorias de *esclarecimento de aspectos metodológicos* e de *organização do relato de pesquisa*.

A distribuição acima enfatiza, em ordem de frequência, uma tendência para:

1. O Estabelecimento de Comparações + ... - com 194 realizações em 222 textos, mais constantemente através do uso de estruturas em Voz Passiva;
2. A Expressão de Objetivos +... - com 142 realizações em 140 textos, principalmente pela adoção de ‘*Noun Phrases + Be +...*’ e de ‘*Verb Phrases (+...)*’;
3. A Indicação de Aspectos Metodológicos – com 107 realizações em 106 textos, especialmente através ‘*Noun Phrases*’;
4. A Topicalização – com 93 realizações em 92 textos, geralmente através de ‘*Noun Phrases*’ e ‘*Noun Phrases*’ + Complementos.

A ênfase em estabelecer algum tipo de mensuração, representada pela primeira categoria (a exemplo de ‘*was associated with a*’, ‘*had no effect on*’ e ‘*a lower risk of*’, dentre outros) aponta um foco das palavras-tópico no desenvolvimento específico do estudo em forma de referências ao seu desenvolvimento como atividade de âmbito experimental especificamente. Dito de outro modo, verifica-se a adoção de um número mais expressivo de associações de palavras que funcionam para indiciar de que modo as investigações se associam, produzem ou não efeito, ou mesmo ocupam lugar como atividades científicas propriamente ditas. Note-se aqui a preponderância da adoção de estruturas passivas (‘*was associated with increased risk*’, ‘*was inversely associated with*’ e ‘*was assessed by using*’,

dentre outros) as quais, de acordo com Hyland (2008^b), representam uma forma importante de destacar relações lógicas nas pesquisas acadêmicas.

A segunda função de maior uso, a expressão dos objetivos / propósitos, por sua vez, aponta uma motivação para a realização dos estudos. O foco desses feixes lexicais é da ordem de reiterar metas e escopos e, portanto, de relacionar a validade das investigações. O trabalho de expressão pragmática se organiza através do uso de sequências recorrentes de palavras tipicamente encontradas na abertura dos *abstracts* desta porção do acervo (tais quais ‘*investigated the effects of*’, ‘*estimated the cost effectiveness*’ e ‘*examine the association between*’). Essas se concretizam, frequentemente, pela presença de palavras-tópico na forma de verbos que apontam o que o estudo realiza (como em ‘*investigate*’, ‘*estimate*’ e ‘*examine*’ acima).

Em contraponto, a próxima função pragmática enfatizada é a indicação de aspectos metodológicos dos estudos (como em ‘*randomized controlled trials rcts*’, ‘*cox proportional hazards models*’ e ‘*case base analysis the*’), cujo objetivo é nomear formas de acordo com as quais os estudos se organizam procedimentalmente. Percebe-se aqui a presença de palavras-tópico que apresentam os arcabouços teóricos dos estudos propriamente. Nesse caso, os dados que evidenciam a função se realizam, principalmente, a partir do uso de ‘*Noun Phrases*’.

Por fim, a quarta ênfase é em topicalização, que se organiza por realizações de linguagem adotadas para definir aspectos temáticos dos estudos (a exemplo de ‘*body mass index bmi*’, ‘*preventive services task force*’ e ‘*quality adjusted life years*’). Nesse caso, é a tônica específica das pesquisas que é referenciada pelas palavras-tópico, organizadas em grupos poli-lexicais, através do emprego de ‘*Noun Phrases*’ e de seus complementos.

Prosseguindo, apresentam-se os pacotes lexicais distribuídos pelo critério léxico-gramatical, para que seja possível notar o modo como a completude do *subcorpus* em questão se organiza em relação à constituição das classes de palavras adotadas pelos autores dos textos. Ao longo desse exercício analítico, novamente, constatou-se que alguns pacotes lexicais poderiam ser relacionados a mais de uma categoria. Entretanto, em se tratando de categorização léxico-gramatical, a abertura das linhas de concordância não potencializa a alteração de classificações, como ocorre em relação às funções retóricas. Para decidir onde agrupar os feixes de palavras, o critério foi por estabelecer associações do modo mais característico em relação à prosa acadêmica. Para dar conta da tarefa, outrossim, a lista de *N-Gram/Clusters* aferida através do AntConc e as linhas de concordância, assim como os

respectivos textos, foram cotejados às classificações de Biber *et al.* (1999) e Biber (2000; 2006). Chegamos então às distribuições que se seguem:

Tabela (10) Ocorrências do Subcorpus Internacional por Classificação Léxico-Gramatical

Categoria Léxico-Gramatical	Frequência	Distribuição	Ocorrência
<i>Noun Phrase</i>	12	12	<i>body mass index bmi</i>
	8	8	<i>preventive services task force</i>
	7	7	<i>and body mass index</i>
	7	7	<i>incremental cost effectiveness ratios</i>
	7	6	<i>quality adjusted life years</i>
	7	7	<i>randomized controlled trials rcts</i>
	6	6	<i>cox proportional hazards models</i>
	6	6	<i>hepatitis c virus hcv</i>
	6	6	<i>reactive oxygen species ros</i>
	6	6	<i>selection randomized controlled trials</i>
	6	6	<i>study selection randomized controlled (fragment)</i>
	5	5	<i>and reference lists study</i>
	5	5	<i>base case analysis the</i>
	5	5	<i>c virus hcv infection</i>

Noun Phrase	Ending in Preposition	6	6	<i>an important role in</i>
		6	6	<i>body mass index in</i>
		5	5	<i>european prospective investigation into</i>
Noun Phrase	+ Prep. Phrase	5	5	<i>prospective investigation into cancer</i>
		6	6	<i>mass index in kg</i>
		5	5	<i>vivo and in vitro</i>
Noun Phrase	+ Embedded Of Phrase Fragment	11	10	<i>the American college of</i>
		10	6	<i>the cost effectiveness of</i>
		8	6	<i>a lower risk of</i>
		8	7	<i>benefits and harms of</i>
		8	8	<i>cochrane central register of</i>
		6	6	<i>homeostasis model assessment of</i>
		6	6	<i>a wide range of</i>
	+ Of Phrase Fragment	12	12	<i>results of base case</i>
		11	11	<i>results of sensitivity analysis</i>
		8	8	<i>central register of controlled</i>
		8	8	<i>the use of a</i>
		8	8	<i>American college of physician</i>
		8	8	<i>register of controlled trials</i>
		7	7	<i>assessment of insulin resistance</i>

		7	7	<i>model assessment of insulin</i>
		7	7	<i>years of follow up</i>
		5	5	<i>aim of this study</i>
		5	5	<i>person years of follow</i>
		5	5	<i>the aim of this</i>
		5	5	<i>and strength of evidence</i>
<i>Noun Phrase</i>	Verb + That Clause Fragment	7	7	<i>these results suggest that</i>
		5	5	<i>these data suggest that</i>
<i>Noun Phrase</i>	+ Be (+...)	13	13	<i>the objective was to</i>
		9	9	<i>this study was to</i>
		7	7	<i>the aim was to</i>
		6	6	<i>the primary outcome was</i>
		5	5	<i>the present study was</i>
		5	5	<i>proportional hazards models were</i>
—	Be + ...	6	6	<i>was to examine the</i>
		5	5	<i>was to investigate the</i>
	Verb Phrase (+...)	13	11	<i>had no effect on</i>
		7	7	<i>assessed risk of bias</i>
		6	6	<i>investigated the effects of</i>
		5	5	<i>to estimate the cost</i>
		5	5	<i>to examine the association</i>

		5	5	<i>plays an important role</i>
		5	5	<i>tested the hypothesis that</i>
		5	5	<i>estimate the cost effectiveness</i>
		5	5	<i>examine the association between</i>
		5	5	<i>has been associated with</i>
		5	5	<i>have been associated with</i>
—	<i>Of Phrase Fragment</i>	12	12	<i>of base case analysis</i>
		9	9	<i>of this study was</i>
		5	5	<i>of bias data synthesis</i>
<i>Prep. Phrase (With Embedded Of Phrase Frag.)</i>		23	16	<i>with the use of</i>
		12	10	<i>on the basis of</i>
		11	10	<i>for the treatment of</i>
		9	6	<i>with a lower risk</i>
		8	7	<i>in this study we</i>
		7	5	<i>in the presence of 1o</i>
		6	6	<i>in vitro and in</i>
		5	5	<i>in the absence of</i>
		5	5	<i>in this review we</i>
		5	5	<i>on the association between</i>
<i>Passive Voice Structure</i>		19	15	<i>was associated with a</i>
		14	10	<i>associated with a lower (fragment)</i>

	12	8	<i>were associated with a</i>
	12	11	<i>were randomly assigned to</i>
	9	7	<i>was associated with higher</i>
	8	6	<i>was associated with lower</i>
	7	7	<i>was not associated with</i>
	7	7	<i>studies are needed to</i>
	7	5	<i>associated with increased risk (fragment)</i>
	7	5	<i>intake was associated with</i>
	6	6	<i>little is known about</i>
	6	6	<i>regression was used to</i>
	6	6	<i>is associated with a</i>
	6	6	<i>trial was registered at</i>
	6	6	<i>was inversely associated with</i>
	6	5	<i>compared with the lowest</i>
	5	5	<i>used to evaluate the</i>
	5	5	<i>were estimated by using</i>
	5	5	<i>was assessed by using</i>
	5	5	<i>this trial was registered</i>

	5	5	<i>models were used to</i>
	5	5	<i>hazards models were used</i>

Fonte: Freitas (2016)

Embora várias categorias tenham sido idênticas àquelas nomeadas nos estudos supra-citados, outras foram acrescentadas com base nos mesmos princípios; são elas ‘*Verb Phrase (+...)*’, ‘*Be +...*’, ‘*Verb Phrase*’, além das classificações que sequenciaram à ‘*Noun Phrase*’ (‘*Ending in Preposition*’, ‘*+ Prepositional Phrase*’, ‘*+ Embedded of Phrase Fragment*’, ‘*+ Of Phrase Fragment*’). Atribuiu-se o fato de ter-se encontrado outras categorias léxico-gramaticais, à questão de que, enquanto *corpora* maiores tendem a gerar um número menor de pacotes, os *corpora* pequenos tendem a gerar mais ocorrências (CHEN e BAKER, 2010). Além disso, há que se observar que esta investigação se debruça sobre produções mais recentes do que os estudos que servem de base para a classificação em pauta. Em outras palavras, não obstante o reconhecimento de uma organização característica na estrutura esquemática do gênero/registo acadêmico, é possível evidenciar que o mesmo não é fixo. Outrossim, aponta-se a prevalência de ‘*Noun Phrases*’ e a adoção de Voz Passiva como tendências desta porção do *corpus*. Especificamente, destaca-se o uso de ‘*Noun Phrases*’ sem complementos e ‘*Noun Phrases*’ seguidos de ‘*Of Phrase Fragments*’. A adoção majoritária da categoria ‘*Noun Phrases*’ confirma um padrão em textos do gênero/registo acadêmico (BIBER *et al.*, 1999; BIBER E GRAY, 2013; HOFFMANN, 2015). O mesmo ocorre em relação ao uso da voz passiva, que desloca o foco de interesse da sentença para o verbo. Conforme referido por Biber e Conrad (2009)¹⁹⁰, o emprego desse tipo de estrutura é muito mais comum em escrita acadêmica do que em outros registros; tendência também verificada por Swales e Feak (2009). Hoffmann (2015), por sua vez, do mesmo modo, reforça a adoção de estruturas passivas como característica do discurso especializado (técnico-científico) (p. 110).

Por fim, cotejam-se os grupos poli-lexicais desta parte do acervo aos 50 feixes lexicais de maior frequência apontados por Hyland (2008^c) a partir da análise de um *corpus* de gêneros/registros acadêmicos do campo de Biologia, compreendendo 794.100 palavras. Os pacotes de 3 e 4 palavras que se seguem, relacionados em ordem de frequência àqueles levantados pelo pesquisador britânico, representam, em realizações no *corpus* internacional,

¹⁹⁰Em estudo citado no cap. 2, seção 2.2.

6.2% do acervo: ‘*a wide range of*’, ‘*in this study we*’, ‘*in the presence of*’, ‘*an important role in*’, ‘*in the absence of*’ e ‘*an importante role*’.

6.2. Do *subcorpus* brasileiro

Deste *subcorpus* com extensão de 90072 palavras, após o trabalho de refinamento, extraíram-se 88 pacotes lexicais com extensão de 4 palavras gráficas. As ocorrências totalizam 764 realizações individuais. A sequência recorrente mais constante foi ‘*of this study was*’ com 37 realizações em 36 textos. São peculiaridades desta porção do acervo a frequência e a distribuição elevadas das primeiras colocações estendidas, aspectos que indiciam alta recorrência de uso de alguns grupos poli-lexicais, como demonstra o quadro que se segue, em que também podem ser notadas as associações de 3 palavras contidas nos pacotes de 4 palavras. Novamente, os cortes (+) apresentados sugerem possibilidades de associações para efeitos ilustrativos apenas. As 10 colocações estendidas mais frequentes ocorreram 233 vezes em 224 textos.

Tabela (11): Frequência e Distribuição das Primeiras Ocorrências do SubCorpus Brasileiro

Sequência	Frequência	Distribuição	Ocorrência
1	37	36	<i>of + this study was</i>
2	36	36	<i>this + study was to</i>
3	30	28	<i>the aim of + this</i>
4	24	23	<i>aim of this + study</i>
5	22	21	<i>as well as + the</i>
6	19	16	<i>in relation to + the</i>
7	19	19	<i>was + to evaluate the</i>
8	17	17	<i>study + was to evaluate</i>
9	16	16	<i>the objective of + this</i>
10	13	12	<i>on the other hand</i> ¹⁹¹

¹⁹¹ Grupo fixo, cuja significado inferencial, não é extraído pela soma das partes, aspecto pouco comum em se tratando de um pacote lexical.

Fonte: Freitas (2016)

Ao realizar o trabalho de distribuição das sequências recorrentes de palavras a partir dos objetivos pragmáticos, chegou-se, novamente, às 7 categorias funcionais postuladas a partir do acervo internacional. Para esta porção dos dados, a distribuição se organizou da seguinte forma:

Tabela (12): Ocorrências do Subcorpus Brasileiro por Classificação Funcional

Categoria Funcional	Frequência	Distribuição	Ocorrência
<i>Estabelecer Comparação /</i>	10	10	<i>is one of the</i>
<i>Associação /</i>	8	7	<i>an increase in the</i>
<i>Possibilidade/</i>	7	7	<i>one of the most</i>
<i>Tecer Avaliação /</i>	6	6	<i>it is possible to</i>
<i>Quantificar</i>	6	6	<i>when compared to the</i>
	5	5	<i>was able to reduce</i>
	5	5	<i>was evaluated using the</i>
	5	5	<i>were not associated with</i>
Apontar Resultados /	12	10	<i>it was observed that</i>
Conclusões /	9	9	<i>these results suggest that</i>
Evidências/	8	7	<i>it was found that</i>
Recomendações /	8	8	<i>results of this study</i>
Consequências/	8	8	<i>the results of this</i>
Indícios dos Dados	8	8	<i>was observed that the</i>
	7	6	<i>it is necessary to</i>
	7	6	<i>was observed in the</i>
	6	6	<i>the results of the</i>
	6	6	<i>results suggest that the</i>
	5	5	<i>it is known that</i>
	5	5	<i>as a result of</i>
	5	5	<i>it was concluded that</i>

	5	5	<i>results indicate that the</i>
	5	5	<i>the effect of the</i>
Expressar Objetivo / Propósito	37	36	<i>of this study was</i>
	36	36	<i>this study was to</i>
	30	28	<i>the aim of this</i>
	24	23	<i>aim of this study</i>
	19	19	<i>was to evaluate the</i>
	17	17	<i>study was to evaluate</i>
	16	16	<i>the objective of this</i>
	13	13	<i>this study aimed to</i>
	12	12	<i>objective of this study</i>
	11	11	<i>the present study was</i>
	11	11	<i>of the present study</i>
	10	9	<i>for the treatment of</i>
	9	9	<i>present study was to</i>
	9	9	<i>of this study is</i>
	8	8	<i>evaluate the effect of</i>
	8	8	<i>to evaluate the effect</i>
	8	8	<i>was to analyze the</i>
	7	7	<i>study was to investigate</i>
	7	7	<i>the purpose of this</i>
	6	6	<i>aimed to evaluate the</i>
	6	5	<i>for the development of</i>
	6	6	<i>study was to analyze</i>
	6	5	<i>the objective was to</i>
	6	6	<i>therefore this study aimed</i>
	6	6	<i>this study is to</i>
	6	6	<i>was to investigate the</i>
5	5	<i>can be used as</i>	

	5	5	<i>of this work was</i>
	5	5	<i>objective of the present</i>
	5	5	<i>study aimed to evaluate</i>
	5	5	<i>the aim of the</i>
	5	5	<i>8the goal of this</i>
	5	5	<i>9this work was to* (sic)</i>
Topicalizar	11	8	<i>the nutritional status of</i>
	9	6	<i>the quality of life</i>
	7	6	<i>the family health strategy</i>
	7	5	<i>quality of life of</i>
	8	8	<i>body mass index bmi</i>
	5	5	<i>mass index bmi and</i>
	5	5	<i>the world health organization</i>
Esclarecer Aspectos Metodológicos	6	5	<i>a cross sectional study</i>
	5	5	<i>cross sectional study with</i>
Organizar o Relato de Pesquisa	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>
	12	12	<i>the analysis of the</i>
	10	6	<i>the end of the</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	7	<i>it was used a* (sic)</i>
	7	6	<i>in this study we</i>
	7	6	<i>the use of the</i>
	6	6	<i>in the city of</i>

	6	6	<i>the other hand the</i>
	6	6	<i>in the process of</i>
	6	6	<i>the beginning of the</i>
	5	5	<i>in the present work* (sic)</i>
	5	5	<i>to the presence of</i>
	5	5	<i>study was carried out</i>
	5	5	<i>were used in the</i>
Contextualizar/Situar	6	6	<i>the university hospital of</i>
	5	5	<i>in children and adolescents</i>

Fonte: Freitas (2016)

Neste caso, a frequência de ocorrências releva ênfase em apenas três categorias, considerando-se as que integram pelo menos 10% das sequências recorrentes de palavras desta porção do *corpus*. Desse modo, aferem-se como tendências do *subcorpus* brasileiro, as seguintes funções retóricas:

1. Expressar Objetivos / Propósitos – com 406 realizações em 362 textos, 51% dessas pelo uso de ‘*Noun Phrases + Be (+..)*’ e de ‘*Noun Phrases + Of Phrase Fragments*’;
2. Organizar o Texto – com 237 realizações em 177 textos, especialmente pelo emprego de ‘*Prepositional Phrases*’;
3. Apontar Resultados / Conclusões + ... – com 118 realizações em 99 textos, desta vez, sem preponderância de padrão léxico-gramatical específico.

A partir da análise das palavras-tópico e seus elementos conexos, tem-se um bloco de textos caracterizado por um alto índice de adoção de colocações estendidas que visam a expressão de objetivos / propósitos das pesquisas. Inclui-se aqui ‘*of this study was*’, o feixe lexical de maior incidência desta parte do acervo, bem como 7 dentre os 10 pacotes lexicais de maior frequência da coleção brasileira, conforme disposto na Tabela 11. Tal prevalência parece apontar uma concentração de interesse dos autores brasileiros em, acima

de tudo, destacar a motivação para realização de seus estudos acadêmicos e, por extensão, para realçar as metas das investigações. Os pacotes lexicais adotados para tanto encontram-se na abertura dos textos e se concretizam principalmente pelo uso de ‘*Noun Phrases*’ e das categorias que se associam a esses (como em ‘*this study was to*’, ‘*the objective of this*’, ‘*the aim of the*’, dentre outros). O que, todavia, se mostra peculiar, a partir da acentuada ênfase em expressão de objetivos e propósitos, é, por extensão, a desênfase nas demais categorias funcionais de associações recorrentes de palavras que indiciam a realização de uma pesquisa científica. Esse aspecto é perceptível mesmo no que tange à diferença de frequência em relação à segunda e terceira categorias de maior uso: os organizadores textuais, que se realizam, tipicamente, através de ‘*Prepositional Phrases*’ (a exemplo de ‘*as well as the*’, ‘*in relation to the*’, ‘*on the other hand*’ e ‘*at the end of*’, entre outros) e as expressões voltadas para apontar resultados/ conclusões/ evidências/ recomendações/ consequências/ indícios dos dados (tais quais ‘*it was observed that*’, ‘*these results suggest that*’, ‘*it was found that*’ e ‘*results of this*’, por exemplo). Acerca dessas duas últimas categorias, enquanto os organizadores textuais representam a classe de menor aproximação temática com a pesquisa relatada *per se*, o apontamento dos resultados / conclusões + é representado por ocorrências que resgatam a dimensão do experimento narrado propriamente, ao cumprirem a função de sinalizar o desfecho das investigações.

Em relação à distribuição léxico-gramatical, além das categorias introduzidas durante o trabalho com a porção internacional dos dados, um grupo classificatório distinto foi encontrado, qual seja ‘*Anticipatory It Pattern*’, apresentando 50 realizações em 46 textos. Revelou-se, outrossim, conforme dispõe a ‘Tabela 12’, uma tendência à concordância com o padrão de textos acadêmicos descrito na literatura (BIBER *et al.*, 1999; HOFFMANN, 2015, dentre outros) pelo uso de ‘*Noun Phrases*’. Esses se realizam, especialmente, seguidos de ‘*Of Phrase Fragments*’, sendo este último grupo aquele cujas associações de palavras apresentam frequência e distribuição mais altas desta porção do acervo, respectivamente 273 vezes em 149 textos. Constatou-se também uma preferência pela adoção de ‘*Prepositional Phrases*’ com 156 realizações, a maior parte desses, conforme já apontado, utilizados para organizar o relato de pesquisa.

Tabela (13): Ocorrências do Subcorpus Brasileiro por Classificação Léxico-Gramatical

Categoria Léxico-Gramatical	Frequência	Distribuição	Ocorrência
<i>Noun Phrase</i>	8	8	<i>body mass index bmi</i>

		7	6	<i>the family health strategy</i>	
		6	5	<i>a cross sectional study</i>	
		5	5	<i>mass index bmi and</i>	
		5	5	<i>the world health organization</i>	
Noun Phrase	Ending in Preposition	5	5	<i>cross sectional study with</i>	
	Ending in Prep Phrase	8	7	<i>an increase in the</i>	
	Embedded Of Phrase Fragment	+	11	8	<i>the nutritional status of</i>
			7	5	<i>quality of life of</i>
			6	6	<i>the university hospital of</i>
	Of Phrase Fragment	+	30	28	<i>the aim of this</i>
			24	23	<i>aim of this study</i>
			16	16	<i>the objective of this</i>
			12	12	<i>objective of this study</i>
			12	12	<i>the analysis of the</i>
			10	10	<i>is one of the</i>
			10	6	<i>the end of the</i>
			9	6	<i>the quality of life</i>
			8	8	<i>results of this study</i>
			8	8	<i>the results of this</i>
			7	7	<i>one of the most</i>
			7	7	<i>the purpose of this</i>
			7	6	<i>the use of the</i>
			6	6	<i>the beginning of the</i>
		6	6	<i>the results of the</i>	
	5	5	<i>objective of the present</i>		
	5	5	<i>the aim of the</i>		

		5	5	<i>the effect of the</i>
		5	5	<i>the goal of this</i>
(Adverb) Noun Phrase	Verb +...	13	13	<i>this study aimed to</i>
		6	6	<i>therefore this study aimed</i>
		5	5	<i>study aimed to evaluate</i>
	Verb + That Clause Fragment	9	9	<i>these results suggest that</i>
		6	6	<i>results suggest that the</i>
		5	5	<i>results indicate that the</i>
	Be (+...)	36	36	<i>this study was to</i>
		17	17	<i>study was to evaluate</i>
		11	11	<i>the present study was</i>
		9	9	<i>present study was to</i>
		7	7	<i>study was to investigate</i>
		6	6	<i>study was to analyze</i>
		6	5	<i>the objective was to</i>
		6	6	<i>this study is to</i>
		5	5	<i>this work was to* (sic)</i>
—	Be + ...	19	19	<i>was to evaluate the</i>
		8	8	<i>was to analyze the</i>
		6	6	<i>was to investigate the</i>
		5	5	<i>was able to reduce</i>
—	Anticipatory It Pattern	12	10	<i>it was observed that</i>
		8	7	<i>it was found that</i>
		7	7	<i>it was used a * (sic)</i>
		7	6	<i>it is necessary to</i>
		6	6	<i>it is possible to</i>

		5	5	<i>it is known that</i>
		5	5	<i>it was concluded that</i>
Verb Phrase (+...)		8	8	<i>evaluate the effect of</i>
		8	8	<i>to evaluate the effect</i>
		6	6	<i>aimed to evaluate the</i>
—	Of Phrase Fragment	37	36	<i>of this study was</i>
		11	11	<i>of the present study</i>
		9	9	<i>of this study is</i>
		5	5	<i>of this work was</i>
Prep. Phrase (With Embedded Of Phrase Frag.)		22	21	<i>as well as the</i>
		19	16	<i>in relation to the</i>
		12	13	<i>on the other hand</i>
		12	7	<i>at the end of</i>
		10	9	<i>for the treatment 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	5	<i>for the development of</i>
		6	6	<i>in the city of</i>
		6	6	<i>in the process of</i>
		6	6	<i>the other hand the (frag)</i>
		5	5	<i>as a result of</i>
		5	5	<i>in children and adolescents</i>
		5	5	<i>in the present work*(sic)</i>
		5	5	<i>to the presence of</i>

<i>Passive Voice Structure</i> (+ Prep. Phrase)	8	8	<i>was observed that the</i>
	7	6	<i>was observed in the</i>
	5	5	<i>study was carried out</i>
	5	5	<i>was evaluated using the</i>
	5	5	<i>were not associated with</i>
	5	5	<i>were used in the</i>
	5	5	<i>can be used as</i>
<i>Other Expressions</i>	6	6	<i>when compared to the</i>

Tabela (10): Corpus CISA – SubCorpus Brasileiro – Pacotes Lexicais: Classificação Léxico-Gramatical

Os grupos poli-lexicais ‘*it was used a*’, ‘*in the present work*’ e ‘*this work was to*’ não são estruturas vernáculas em língua inglesa, em oposição à ‘*a... was used*’, ‘*in the present paper/research*’ e ‘*this paper/research was to*’. De todo modo, o emprego desses feixes lexicais mostrou-se representativo a ponto de constituírem pacotes lexicais com 17 realizações em 17 textos deste segmento do *corpus*. Ainda, ‘*in the present study*’, de acordo com Hyland (2008^c) está entre as colocações estendidas mais frequentes da prosa acadêmica. Em função desses aspectos, optou-se por manter tais ocorrências nesta parte do acervo, quando do trabalho de refinamento dos pacotes lexicais. Além disso, indiciam a necessidade de atenção por parte dos autores brasileiros tanto em relação à questão gramatical, quanto a peculiaridades de uso.

Ao comparar esta porção do acervo aos 50 pacotes lexicais mais frequentes da prosa acadêmica no campo de Biologia, a partir da descrição de Hyland (2008^c), relacionam-se as seguintes ocorrências de 3 e 4 palavras em ordem de frequência : ‘*as well as the*’, ‘*on the other hand*’ (de ‘*on the other hand*’ e de ‘*the other hand the*’), ‘*at the end of*’, ‘*is one of the*’, ‘*the end of the*’, ‘*the present study*’ (de ‘*of the present study*’ e ‘*the present study was*’), ‘*in the present study*’ (de ‘*in the present work*’), ‘*in the presence of*’, ‘*the quality of*’, ‘*it was found that*’, ‘*the results of*’, ‘*in this study*’ (de ‘*in this study the*’ e ‘*in this study we*’), ‘*the beginning of the*’, ‘*the results of the*’ e ‘*as a result of*’. Essas representam 20.4% do subcorpus brasileiro.

Em segmento, a fim de destacar padrões que tipificam a porção nacional do acervo, apresenta-se a constituição das associações recorrentes de palavras que integram as coleções de *abstracts* que formam o *subcorpus* brasileiro:

1. Tabela (14): Ocorrências do Subcorpus Brasileiro de Publicações

Sequência	Frequência	Distribuição	Ocorrência
1	12	12 objetivos	<i>of this study was</i>
2	12	12 objetivos	<i>this study was to</i>
3	8	8 objetivos	<i>was to evaluate the</i>
4	6	6 objetivos	<i>study was to evaluate</i>
5	6	6 objetivos	<i>the aim of this</i>
6	6	6 objetivos	<i>the objective of this</i>
7	5	5 objetivos	<i>aim of this study</i>

Fonte: Freitas (2016)

Este *subcorpus* representa 32.9% dos textos do acervo brasileiro e é composto por 118 *abstracts*, 19170 palavras e 3852 tipos. A distribuição das ocorrências iniciada pelas palavras-tópico e respectivos elementos conexos é homogênea: todos os feixes lexicais cumprem a função de expressar objetivo/propósito da pesquisa descrita. Os pacotes representam 55 realizações em um número idêntico de textos. A composição léxico-gramatical se organiza por ‘*Noun Phrases + Of Phrase Fragments*’ (‘*the aim of this*’, ‘*the objective of this*’ e ‘*aim of this study*’), por ‘*Noun Phrases + Be +...*’ (‘*this study was to*’ e ‘*study was to evaluate*’), por ‘*Of Phrase Fragments*’ (‘*of this study was*’), bem como por ‘*Be +...*’ (‘*was to evaluate the*’). A tendência, por conseguinte, aponta o uso de *Noun Phrases* (35 realizações), para apontar as finalidades de realização das publicações brasileiras.

2. Tabela (15): Ocorrências do Subcorpus Brasileiro de Teses e Dissertações

Sequência	Frequência	Distribuição	Ocorrência
1	18	18 objetivo	<i>of this study was</i>
2	18	18 objetivo	<i>this study was to</i>

3	17	17 objetivo	<i>the aim of this</i>
4	14	14 objetivo	<i>aim of this study</i>
5	14	14 organizar	<i>as well as the</i>
6	11	8 organizar	<i>in relation to the</i>
7	9	9 objetivo	<i>of the present study</i>
8	9	8 organizar	<i>on the other hand</i>
9	8	7 organizar	<i>in the presence of</i>
10	8	8 objetivo	<i>the present study was</i>
11	8	8 objetivo	<i>this study aimed to</i>
12	8	8 objetivo	<i>was to evaluate the</i>
13	7	6 objetivo	<i>for the treatment of</i>
14	7	5 apontar	<i>it was observed that</i>
15	7	7 objetivo	<i>study was to evaluate</i>
16	7	7 apontar	<i>these results suggest that</i>
17	6	5 objetivo	<i>for the development of</i>
18	6	6 objetivo	<i>the present study was to</i>
19	6	6 organizar	<i>the analysis of the</i>
20	5	5 organizar	<i>in this study the</i>
21	5	5 comparar	<i>is one of the</i>
22	5	5 objetivo	<i>the objective of this study</i>
23	5	5 comparar	<i>one of the most</i>
24	5	5 objetivo	<i>study was to investigate</i>
25	5	5 objetivo	<i>the objective of this</i>
26	5	5 organizar	<i>the other hand the</i>

Fonte: Freitas (2016)

O subcorpus organizado a partir de teses e dissertações representa a maior porção dos dados brasileiros: 39.1% do acervo, na forma de 140 textos que totalizam 46147

palavras, distribuídas em 6299 tipos e 26 pacotes lexicais. Os mesmos se dispersam em 223 realizações ao longo de 214 textos. A principal função das associações de palavras, novamente, é a expressão dos objetivos/propósitos dos estudos que descrevem (*'this study was to'*, *'aim of this study'*, *'the presente study was to'*, etc.) com 141 realizações em 139 textos, especialmente por meio de *'Noun Phrases (+...)'* e/ou de *'Of Phrase Fragments'*. A segunda função de maior uso é a organização do relato de pesquisa (*'as well as the'*, *'in the presence of'*, *'on the other hand'*, etc.) com 112 realizações em 107 textos, principalmente pelo emprego de *'Prepositional Phrases'*. Uma vez que as demais funções encontradas representam apenas, respectivamente, 6.2% e 4.4% do *subcorpus*, ficando abaixo da linha de corte de 10%, replica-se aqui o padrão das duas principais funções relacionadas nos dados brasileiros como um todo.

3. Tabela (16): Ocorrências do Subcorpus Brasileiro de Monografias

Sequência	Frequência	Distribuição	Ocorrência
1	7	6 objetivo	<i>of this study was</i>
2	7	5 objetivo	<i>the aim of this</i>
3	6	5 organizar	<i>as well as the</i>
4	6	6 organizar	<i>the analysis of the</i>
5	6	6 objetivo	<i>this study was to</i>
6	5	5 organizar	<i>in relation to the</i>
7	5	5 objetivo	<i>the objective of this</i>

Fonte: Freitas 2016

Por fim, a porção do acervo que contém os *abstracts* de monografias é composta por 100 textos (27.9% da coleção brasileira), 24755 palavras e 4399 tipos, os quais perfazem 7 pacotes lexicais, distribuídos em 42 realizações ao longo de 38 textos. Novamente, os pacotes se organizam, respectivamente, em expressão de objetivos/ propósitos e organização do relato de pesquisa. O nexó léxico-gramatical, se classifica pela adoção de *'Noun Phrases + ...'*, pelo uso de *'Of Phrase Fragments'* e de *'Prepositional Phrases'*.

Em resumo, analisados separadamente, os *subcorpora* que integram o acervo brasileiro reiteram as funções de expressão de objetivos / propósitos e de organização textual, entretanto, não reafirmam a função retórica de apotamento de resultados / conclusões das investigações, como o fazem quando analisados conjuntamente. Portanto, a partir da unidade de análise, os textos salientam uma tendência, primeiramente, para o apontamento dos objetivos dos experimentos e, em segundo lugar, para elucidar o modo como os relatos se engendram. O emprego de ‘*Noun Phases*’ e de ‘*Of Phrase Fragments*’ como padrão léxico-gramatical de destaque, por seu turno, foi frequente tanto nas partes, quanto no todo.

Para finalizar a apresentação do *corpus*, organizou-se a breve síntese que se segue, cujas principais implicações comparativas serão desdobradas no próximo capítulo:

- a) Os feixes lexicais extraídos do *Corpus* CISA confirmam o princípio segundo o qual pacotes lexicais de 3 palavras são muito frequentes e geralmente se abrigam em sequências de 4 palavras, a exemplo de ‘*had + no affect on*’ e ‘*as well as + the*’;
- b) Os dois *subcorpora* principais (internacional e brasileiro) se distinguem em termos de frequência e distribuição das principais ocorrências;
- c) Das 7 categorias de ordem pragmática estabelecidas a partir das afinidades entre os dados, a coleção internacional apresenta 4 funções principais, enquanto a coleção brasileira apresenta 3. A divisão, tanto das principais categorias, quanto das demais, comporta diferenças da ordem de distribuição e frequência, o que indicia distinções retóricas entre os dois *subcorpora*;
- d) Apesar da existência de um padrão característico para o gênero/registo acadêmico descrito a partir de associações léxico-gramaticais típicas (BIBER *et al.*;1999 e BIBER, 2000; 2006), a ocorrência de outras configurações em nosso *corpus* evidencia que o discurso científico não é fixo;
- e) O uso majoritário de ‘*Noun Phrases*’ aproxima as distribuições léxico-gramaticais de ambas as porções do acervo, apesar da ocorrência de desigualdades estruturais relacionadas a outras constituições de feixes de palavras;
- f) A presença de feixes lexicais não vernaculares, dentre esses in ‘*the present work*’ (sic) (‘*in the present study*’), descrita como uma das colocações estendidas mais frequentes da prosa acadêmica (HYLAND, 2008^c), indicia que há aspectos de ordem léxico-

gramatical e peculiaridades de uso a serem observados na produção acadêmica de autores brasileiros;

- g) Cotejando-se o *Corpus CISA* aos grupos poli-lexicais descritos por Hyland (2008^c), em amplo estudo a respeito de expressões típicas das ciências biológicas, o *subcorpus* internacional apresenta 6.2% desses pacotes lexicais e o *subcorpus* brasileiro 20.4%.

Neste capítulo, apresentaram-se as duas porções maiores de dados que compõem o *Corpus CISA*, as coleções que integram a parte brasileira do acervo e as respectivas classificações, assim como teceu-se uma breve síntese dos padrões aferidos. Na próxima parte, estabelece-se uma análise comparativa detalhada dos resultados.

7. CAPÍTULO 7

“There is more to get out of corpora than just words.”

Stephan Gries (2009, p. 9)

‘THESE RESULTS SUGGEST THAT...’: SÍNTESE DOS RESULTADOS E RETOMADA DAS QUESTÕES DE PESQUISA

Inicia-se este capítulo pelo tratamento das questões de pesquisa concebidas para a investigação, quais sejam:

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?

Vale lembrar 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. Dito de outro modo, faz-se referência à constituição, propriamente, de nossa unidade investigativa, a exemplo da tabela que se segue:

Tabela (17): Palavras-Tópico e Elementos Conexos

Pacotes Lexicais	Palavras-Tópico	Elementos Conexos
<i>with the use of</i>	<i>use</i>	<i>with the ... of</i>
<i>was associated with a</i>	<i>associated</i>	<i>was... with a</i>
<i>had no affect on</i>	<i>effect</i>	<i>had no ...on</i>
<i>the objective was to</i>	<i>objective</i>	<i>the...was to</i>
<i>body mass index bmi</i>	<i>body mass index</i>	<i>... bmi</i>
<i>of this study was</i>	<i>study</i>	<i>of this...was</i>
<i>aim of this study</i>	<i>aim / study</i>	<i>...of this...</i>
<i>in relation to the</i>	<i>relation</i>	<i>in...to the</i>
<i>was to evaluate the</i>	<i>evaluate</i>	<i>was to... the</i>

<i>on the other hand</i>	<i>on the other hand</i>	—
--------------------------	--------------------------	---

Freitas (2016)

Para lidar com os questionamentos estipulados, estabeleceu-se um cotejamento das duas porções maiores que formam o acervo de dados. O propósito desta parte do texto é de, a partir de um contraste entre os resultados, destacar os modos como a unidade analítica do estudo confere padrões aos textos do *corpus*, apesar de ser constituída de porções fragmentárias do léxico.

7.1. Do tipo de variabilidade lexical e do comportamento estrutural e funcional dos dados que integram o *corpus*

Ao co-relacionarmos os dois *subcorpora* que compõem os dados do estudo, é possível perceber que, embora as porções internacional e a brasileira sejam de proporções quase idênticas, respectivamente 90.098 palavras e 90.072 palavras, o acervo de origem nacional, mesmo com menos pacotes lexicais (96 x 88), caracteriza-se por um maior número de realizações de ocorrências (672 x 764). Em outras palavras, ocorre maior repetição de grupos poli-lexicais no *subcorpus* brasileiro, o que indicia uma menor variação na adoção de palavras- tópico e dos respectivos elementos conexos. A exemplo disso, as 10 ocorrências de maior frequência da coleção internacional atingem 142 efetivações. Na coleção brasileira, as principais ocorrências atingem 233 realizações, como é possível constatar no quadro que se segue, em que os pacotes lexicais desse *subcorpus* apresentam frequências mais altas do que os correspondentes em ordem de constância nos dados internacionais:

Tabela (18): Ocorrências de Maior Frequência do *Corpus* CISA

<i>Subcorpus</i> Internacional			<i>Subcorpus</i> 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>

Freitas (2016)

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

1. A dispersão dos pacotes ao longo do subcorpus 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. O Corpus CISA corrobora a constatação de Hyland (2008^b) de que escritores menos experientes apoiam-se 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. No caso desta investigação, entretanto, a constatação se faz evidente por conta de maior recorrência relacionada a uma menor variedade lexical;
2. Em termos interpretativos, confirma-se 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.

Tendo em vista os desdobramentos expostos, cabe discutir, então, se um maior apoio em linguagem formulaica seria, necessariamente, um demérito dos textos brasileiros em relação ao padrão especializado. Para tanto, é oportuno lembrar que, independentemente de

os dados corroborarem o nexa aferido por Hyland (2008^b) mencionado, 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 noviços ou usuários menos fluentes (DE COCK, 2000, CORTES, 2002, HYLAND, 2008^a, 2008^b, 2008^c, SALAZAR, 2011).

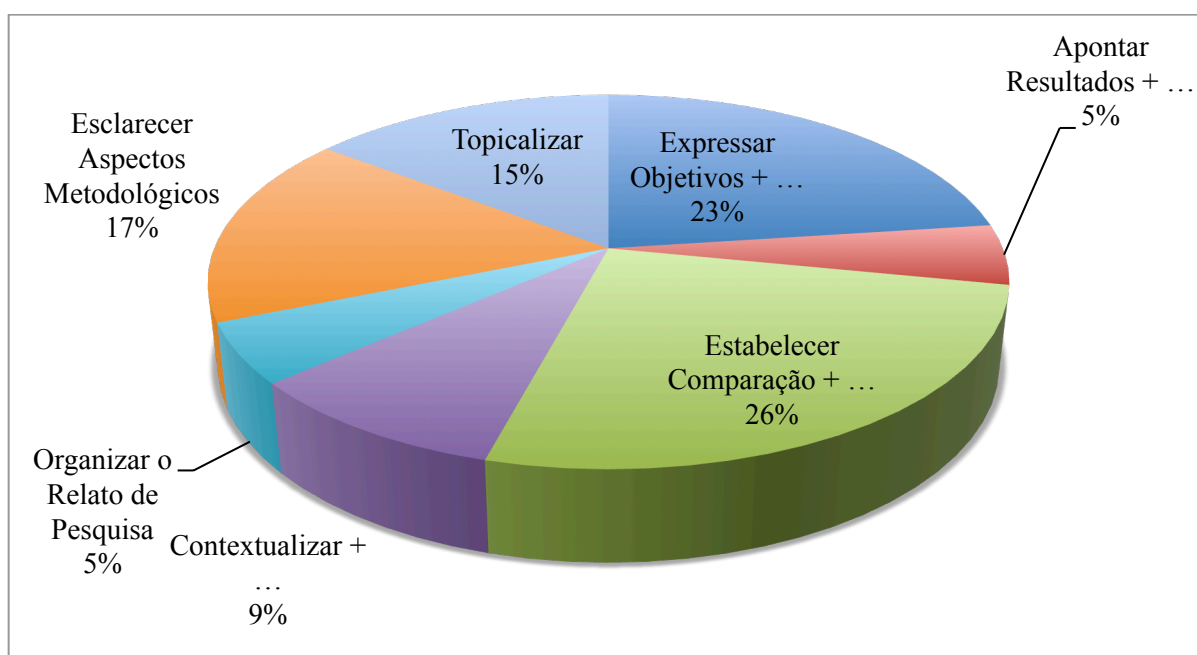
Com base no acervo, observa-se que um maior apoio em linguagem formulaica não é, necessariamente, deflagrador de textos fora do padrão característico de um domínio especializado, uma vez que tal linguagem seja capaz de refletir os modos de expressão e, com isso, os modelos cognitivos e culturais adotados pelos especialistas em um tema. Para ratificar esse achado, vale lembrar um aspecto mencionado no capítulo anterior, de que as palavras-tópico e elementos conexos do segmento brasileiro do *corpus* representam, em termos de ocorrência individuais, 20.4% do acervo dos 50 pacotes lexicais mais frequentes da prosa acadêmica no campo biológico, descrito por Hyland (2008^c) em extensa pesquisa. O acervo internacional do *Corpus CISA*, por seu turno, representa 6.2% dessas colocações estendidas de alta frequência. Ainda, se considerarmos que, nesta pesquisa, o *subcorpus* brasileiro é fruto de produções de pesquisadores em diferentes graus de *expertise* acadêmica, podemos nos alinhar a outra constatação de Hyland (2008^b), de que a adoção de um maior número de associações recorrentes de palavras é um modo de os noviços reforçarem seu pertencimento à comunidade acadêmica. Poderíamos, assim, refletir que, uma vez que os pesquisadores brasileiros se acham fora do eixo de produção intelectual onde se encontra a comunidade de maior prestígio em termos de publicações científicas do campo da saúde, a um maior apoio em associações recorrentes de palavras pode ser atribuído outro modo de enfatizar uma relação de pertencimento ao grupo internacional de pesquisa. Ademais, concordo com Cortes (2002), quando a autora infere 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. Enfim, os redatores brasileiros usam, repetidamente, as mesmas estruturas e, com isso, perfazem um repertório pouco variado.

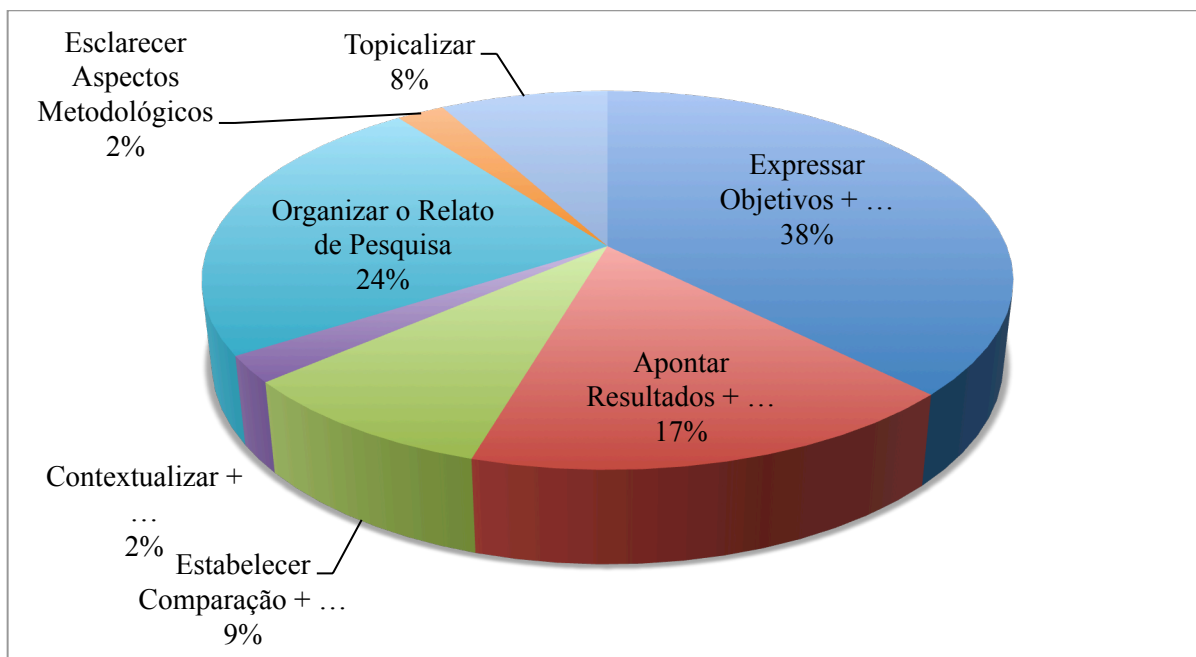
Além das diferenças apontadas em relação à saliência no emprego dos feixes de palavras, os acervos também se caracterizam pela presença de desigualdades funcionais e estruturais. Tais desigualdades sugerem outras distinções com relação, tanto ao modo de produzir um relato científico, quanto ao uso da língua inglesa. Para ilustrar esse aspecto, apresentam-se os perfis gráficos que condesam os dados:

Figura (17) Gráfico da Distribuição Funcional do Subcorpus Internacional



Freitas (2016)

Figura (18) Gráfico da Distribuição Funcional do Subcorpus Brasileiro



Freitas (2016)

Ao compararmos as distribuições 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 representam 81% das ocorrências que, nesta etapa da exposição, para fins de síntese analítica, compilamos 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'*¹⁹², 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

¹⁹² Destaque-se a alta frequência dessas quatro expressões, que juntas representam 15.5% das realizações do subcorpus.

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 de propósitos, 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. Outrossim, na porção internacional dos dados, 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. Destacam-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'*¹⁹³) 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.

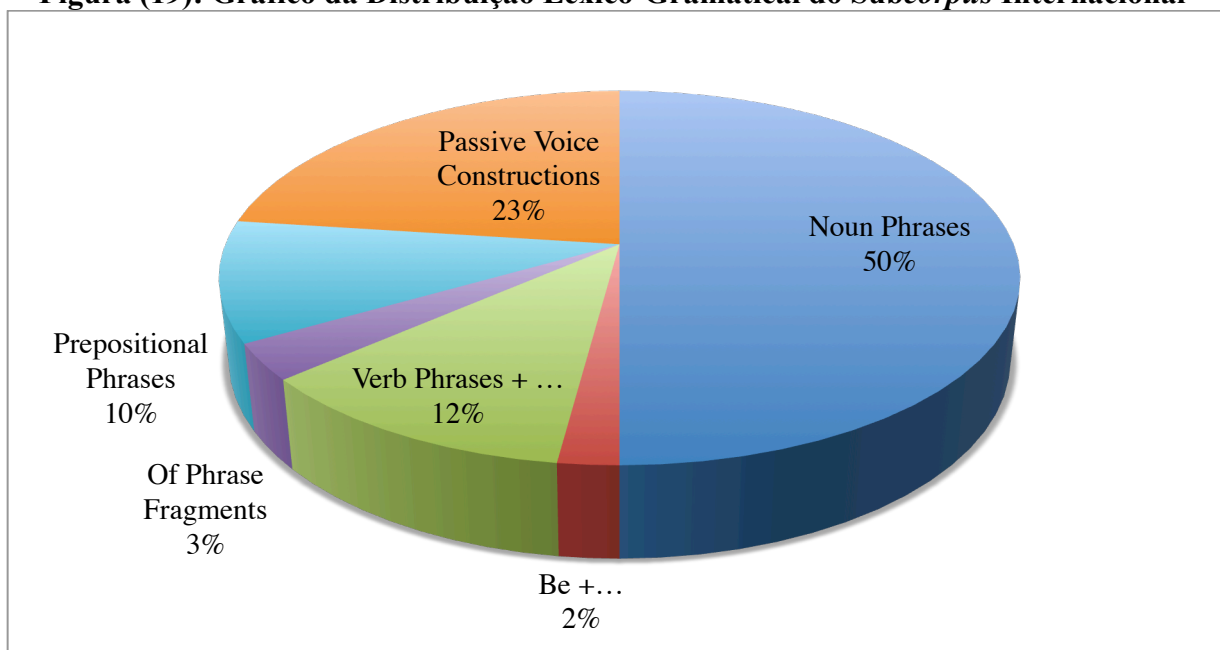
¹⁹³ Esses quatro pacotes lexicais listados encontram-se no quadro das ocorrências mais frequentes do *corpus*, apresentado nas pág. 185-6. Os mesmos representam 16.6% das realizações dos dados brasileiros.

Em resumo, ao adotarmos 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*, distribuídos com base nas palavras-tópico e respectivos elementos conexos, 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 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. Poderia estar expressa aqui a preocupação de parte dos autores dos textos brasileiros com o fortalecimento de seu *status* investigativo em prol do reconhecimento dos membros especialistas da comunidade acadêmica, aludida anteriormente. De acordo com Swales (1990), “essa razão formata a estrutura esquemática do discurso e influencia e delimita a escolha de conteúdo e estilo” (p. 58). Está-se apontando, portanto, que o grupo de pesquisadores brasileiros do acervo, com base na agenda de ocupar um nicho no espaço investigativo mais amplo, estabeleceu objetivos comunicativos relacionados a seu contexto específico de atuação, os quais norteiam o estilo e a estrutura dos textos do *subcorpus* brasileiro.

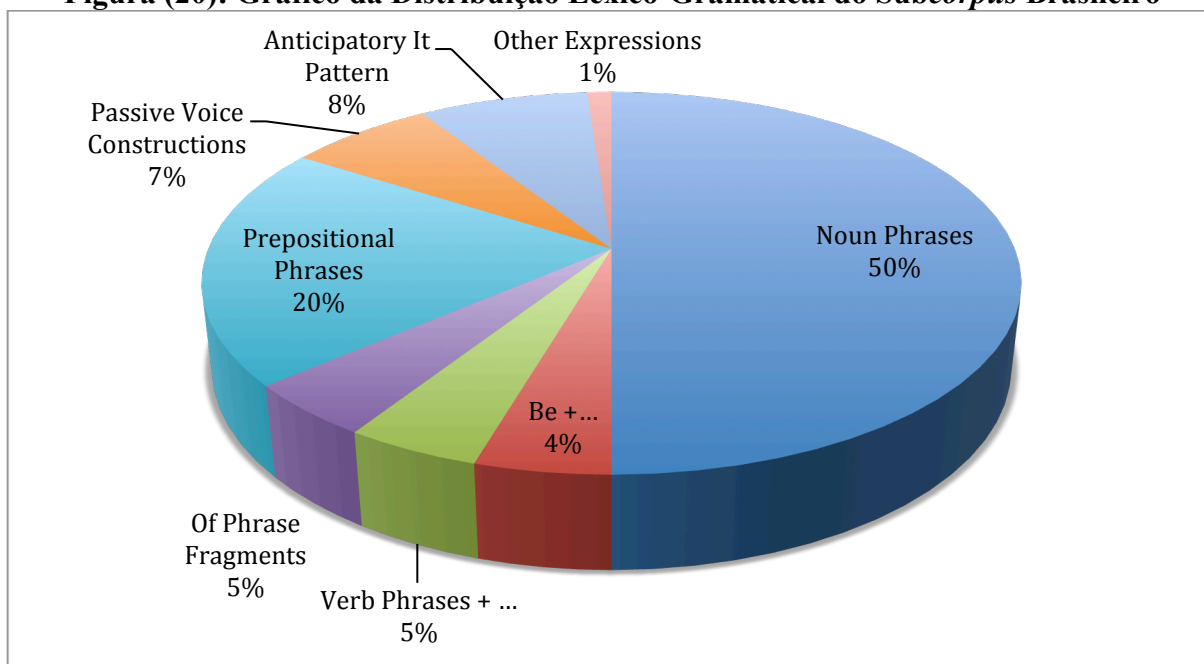
Em sequência, apresentam-se as distribuições dos perfis léxico-gramaticais e, com isso, relacionam-se os modos como a segunda classificação dos feixes de palavras reforça a primeira:

Figura (19): Gráfico da Distribuição Léxico-Gramatical do Subcorpus Internacional



Freitas (2016)

Figura (20): Gráfico da Distribuição Léxico-Gramatical do Subcorpus Brasileiro



Freitas (2016)

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*’¹⁹⁴, ‘*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.

Primeiramente, ressaltam-se as categorias de destaque em comum 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, sessenta por cento 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 vem a ser a segunda categoria de maior frequência nos dados de origem nacional, 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

¹⁹⁴ Ocorrência de maior saliência do *subcorpus*.

objetiva do que o estilo discursivo acadêmico anglo-americano. Para ilustrar esse ponto, retomemos os quadros comparativos que correlacionam o emprego de locuções preposicionadas à função de organização do relato de pesquisa nas duas porções do acervo:

Tabela (19): 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>
Subcorpus Brasileiro	Frequência	Distribuição	Ocorrência
Organizar o Relato de Pesquisa	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>

Freitas (2016)

Com base na tabela cima, tem-se 131 realizações da estrutura no acervo brasileiro em relação a 47 nos dados internacionais como evidência, portanto, de uma retórica mais prolixa presente 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. Vale notar, portanto, 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 nos 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.

Discutidas as principais construções dos dois acervos, vale ressaltar, igualmente, que a porção brasileira inclui a estrutura ‘*anticipatory it pattern*’, distinta das demais categorias da contraparte internacional. A exemplo dessa, registramos ‘*it was observed that*’, ‘*it was found that*’, ‘*it is known that*’ e ‘*it was concluded that*’. O emprego desses feixes lexicais centrados em palavras-tópico na forma de verbos é descrito na literatura de escrita acadêmica (DAYRELL, 2009; GROOM, 2005) como forma de modular o discurso¹⁹⁵. Para Dayrell (*op. cit.*) e Groom (*op. cit.*), no uso de tais expressões está implicado, por extensão, o apagamento da autoria de uma asserção e, com isso, menor agência

¹⁹⁵ *Hedging propositions.*

de parte dos escritores que os adotam. No entanto, os dados a partir do *Corpus CISA* não corroboram essa descrição, pois a maior parte dessas colocações estendidas serve ao apontamento de resultados +... Desse modo, uma vez que o propósito da função retórica em questão é de fazer referência ao fazer científico propriamente dito, defende-se que o uso de palavras-tópico e elementos conexos como moduladores do discurso, nesse caso, conduz o foco de interesse do leitor para a compreensão do estudo, a exemplo de “*It was observed that the intestinal injury caused by enterotomy and intestinal obstruction raise the quantities of MMPs in the hoof laminae*”.

7.2. Da síntese dos resultados

Com base no cotejamento entre as duas porções maiores que integram o *corpus*, destaca-se o modo como a unidade analítica do estudo, constituída a partir de porções fragmentárias do léxico, confere padrões aos textos do *corpus*. De modo conciso, com base nos aspectos de frequência e variabilidade lexical descritos, registram-se os seguintes traços e desdobramentos acerca do comportamento funcional e estrutural do acervo:

1. O *subcorpus* brasileiro se caracteriza pela saliência ou maior repetição de grupos poli-lexicais. A dispersão dos pacotes ao longo do *subcorpus* demonstra que os autores adotam a mesma construção repetidamente, o que revela uma maior convencionalidade dos textos;
2. O maior montante de realizações aponta um percentual mais alto de ocorrências nos dados brasileiros, embora, em números absolutos, o *subcorpus* internacional contenha um número mais elevado de pacotes lexicais;
3. Um maior apoio em linguagem formulaica não é, necessariamente, deflagrador de textos fora do padrão característico de um domínio especializado, conquanto o padrão adotado reflita os modos de expressão e, com isso, os modelos cognitivos e culturais adotados pelos especialistas em um tema;

4. A adoção de um maior número de realizações de pacotes lexicais dá mostras de ser um modo de os pesquisadores brasileiros reforçarem seu pertencimento à comunidade acadêmica do hemisfério norte, em função do emprego recorrente de construções que, em tese, tipificam a retórica acadêmica anglófona;
5. O *subcorpus* brasileiro se mostra menos característico do padrão de expressão do conhecimento especializado dos textos em língua inglesa contidos no acervo internacional, tendo em vista a baixa variação das principais colocações estendidas;
6. As associações entre palavras-tópico e elementos conexos relevam que as distribuições dos perfis retóricos e léxico-gramaticais dos dados corroboram-se mutuamente;
7. Embora a prevalência de construções de base nominal que abriga as palavras-tópico aproxime e tipifique os dois *subcorpora* como representantes do registro acadêmico, outras desigualdades estruturais, além de diferenças funcionais entre os acervos, sugerem distinções no emprego dos grupos poli-lexicais, com relação ao modo de produzir um relato científico;
8. As principais associações de palavras-tópico e elementos conexos integrantes do acervo brasileiro relevam maior saliência também em relação às funções referenciais aferidas;

9. 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;
10. Com base na agenda de ocupar um nicho no espaço investigativo mais amplo, o grupo de pesquisadores brasileiros estabeleceu objetivos comunicativos próprios que norteiam o estilo e a estrutura dos textos e, com isso, impactam o registro acadêmico produzido;
11. Os dados brasileiros sugerem a presença de uma retórica menos objetiva do que o estilo discursivo acadêmico anglo-americano, com base no percentual de uso de locuções preposicionadas, para cumprir a função de detalhar aspectos da organização das investigações;
12. A maior adoção de voz passiva nos dados internacionais indicia maior articulação textual em promover agência ao objeto do fazer científico, como pode ser particularmente relevante para a retórica acadêmica enfatizar;
13. As associações de palavras-tópico e elementos conexos iniciadas por verbos são mais frequentes no acervo internacional, em que expressam tanto a função de apontar a motivação para desenvolvimento dos estudos, quanto referência ao fazer científico em si;
14. Nos dados brasileiros, o uso de *hedges* ou moduladores do discurso, através do emprego da estrutura '*anticipatory it pattern*', aponta para a condução do foco de interesse do leitor em direção à compreensão do estudo, em referência ao fazer científico propriamente dito.

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 pacotes lexicais para expressar a finalidade de justificar a realização do trabalho acadêmico, empreitada que, neste segundo conjunto de dados, 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.

Concluída a análise comparativa dos dados e respondidas as questões de pesquisa, cabe retomar a epígrafe que abre este capítulo da tese, segundo a qual “há mais do que palavras para se extrair de *corpora*” (GRIES, 2009, p. 9)¹⁹⁶. Dessa asserção, se poderia depreender que os *corpora* geram evidências valiosas a respeito do uso da linguagem e que tais evidências contêm rica aplicabilidade didática. 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. Especificamente, não se trata de homogeneizar a redação de *abstracts* brasileira no campo da saúde para que seja idêntica ao padrão anglo-americano. 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 ao acadêmico 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, é preciso orientar a prática linguística para comportar a 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

¹⁹⁶“*There is more to get out of corpora than just words*” (GRIES, 2009, p. 9).

linguísticas informadas. Com isso, é possível desenvolver referências acerca do tipo de linguagem que se espera que um pesquisador produza. Para dar conta desse postulado, professor e estudante precisam dispor de dados de reconhecimento dos diferentes modos de dizer e princípios de combinabilidade de elementos linguísticos dos *abstracts* do campo da saúde. Não obstante, 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.

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*.

Isso posto, passamos ao capítulo final do estudo, em que se tecem considerações pedagógicas, acerca do ensino e do lugar das associações recorrentes de palavras em face à produção de *abstracts*.

8. CAPÍTULO 8

“We must apprehend language events in their contexts as shaped by the creative acts of speaking persons, decode scientific discourse and make the discourse accessible in education.”

Firth (1957, p.190)

‘STUDIES ARE NEEDED TO...’

Cabe agora retomar o roteiro percorrido ao longo deste trabalho, até este capítulo final. Iniciamos a pesquisa discutindo a noção de *texto*, concebido como linguagem em uso (HALLIDAY e HASAN, 1978), pela exploração de uma revisão da trajetória dos estudos textuais (HOFFMANN, 2015; BLUHDORN e ANDRADE, 2005; BRINKER, 2005; CIAPUSCIO, 2003; KOCH, 2002; BEAUGRANDE e DRESSLER, 1981; VAN DIJK, 1979; 1980). Em segmento, com base nas ideias de Bakhtin (1997), Swales (1990) e Biber (1995), situamos o modo com que os textos funcionam como modelos de instâncias comunicativas e, com isso, organizam-se em *gêneros discursivos*. Ao fazê-lo, analisamos a perspectiva de *registro* contida na concepção biberiana, como um ponto de confluência entre os três pesquisadores nomeados. Isso porque, lembrando a discussão já promovida, Biber (*op. cit.*) estabelece a ideia de sistematicidade que integra os *tipos relativamente estáveis*, de Bakhtin (*op. cit.*), à *estrutura esquemática do discurso reconhecida pelos membros de uma comunidade*, de Swales (*op. cit.*). Também porque aquele autor agrega uma preocupação com a frequência de ocorrência dos itens lexicais aos estudos da temática dos textos acadêmicos.

Em segmento, abordamos a constituição do *texto especializado*, mapeando as características através das quais esse se torna instrumento de nossa participação no mundo do conhecimento científico. Para isso, nos pautamos, especialmente, pelas concepções de cunho linguístico e discursivo de Hoffmann (1998; 2015), que nos faz ir além das terminologias mais peculiares dos textos “especializados” de diferentes domínios. Em sequência, historiamos o cenário de surgimento do texto concebido como uma virada na história dos manuscritos acadêmicos, o *abstract*, as características e os principais estudos no tema (CILVETTE e PÉREZ, 1986; DAYRELL, 2015; GLEDHILL, 2015; 2012; GRAETZ, 1985; SALAGER-MEYER, 1991; SWALES, 1990; SWALES e FEAK, 2009, dentre outros). Posteriormente, exploramos a *LC* (MCENERY e WILSON, 1996; SINCLAIR, 1997), forma de aproximação da linguagem eleita para a investigação, as pesquisas em *EAP* (BIBER e CONRAD, 2009) e a unidade analítica do estudo, os *pacotes lexicais* (BIBER *et al.*, 1999), em meio à constituição dos *abstracts*. As questões concebidas para a pesquisa foram então explicitadas, assim como o trajeto metodológico da exploração empírica e o *corpus* do trabalho. Feito isso, os pacotes lexicais extraídos dos dados foram apresentados e apontaram-se as diferenças e semelhanças entre as partes do acervo reunido. Ao estabelecermos a comparação analítica, revelou-se de que modo os fragmentos do léxico, constituídos pelas

associações de palavras recorrentes que formam os *lexical bundles* encontrados, contribuem para dar forma ao perfil dos *subcorpora* que compõem o acervo.

Por fim, neste capítulo em que concluímos o estudo, firma-se a ponte com o ensino de escrita acadêmica. Com isso, reforça-se o objetivo geral de contribuir para o aprendizado da produção de *abstracts* em língua inglesa no campo das Ciências da Saúde, assim como de apoiar a formação educadores linguísticos e os avanços do campo de EAP. Além disso, projetam-se perspectivas de continuidade para o estudo e reflete-se sobre o alcance da pesquisa.

8.1. Do modo como os fragmentos do léxico conferem padrões aos *abstracts*

São dois os fundamentos postulados para construir uma base para o ensino/aprendizagem da redação fluente de *abstracts* no campo pesquisado:

1. A noção de que nada na comunicação humana escapa a algum tipo de gênero discursivo e textual e que, portanto, o todo é ponto de partida para dar sentido às partes;
2. O princípio idiomático de Sinclair (1987), de acordo com o qual “um usuário da língua tem a sua disposição um grande número de frases semi-pré-construídas, que constituem escolhas únicas, mesmo que elas pareçam ser analisáveis em segmentos” (p. 120).

Para consolidar o trabalho pedagógico, estabelecem-se quatro etapas, relacionadas aos fundamentos elencados:

1. A Ênfase nos Objetivos e Contextos;
2. A Leitura Crítica;
3. O Trabalho com Pacotes Lexicais;
4. A Prática de Redigir o Texto.

1. Como ponto de partida, para enfatizar a relevância do *abstract* como texto

acadêmico, propõe-se a tarefa de que se elenquem, a partir das experiências ou leituras de mundo dos participantes, objetivos e contextos de uso do texto em língua inglesa, bem como os principais desafios em sua produção. Pode-se contrastar ao texto produzido em língua portuguesa inclusive. Desse modo, o exercício de situar o gênero e sua escrita, ao mesmo tempo que estranha o objeto em função da reflexão promovida, aproxima-o do aprendiz, pela busca de situações de adoção em cenários de conhecimento do usuário e pelo compartilhamento de experiências. Com essa tarefa, tem-se por meta pontuar o motivo pelo qual o *abstract* se apresenta como o cartão de visitas dos textos acadêmicos e por que sua escrita é complexa, dadas a densidade lexical e as diferenças estruturais entre as duas línguas. Como subsídio teórico para organizar uma apresentação de conclusão desta primeira parte, sugere-se a adoção do conteúdo do capítulo 3: ‘*Abstracts*, uma Virada na História dos Textos Acadêmicos’.

2. Em segmento, com sustentação no primeiro fundamento apresentado acima - segundo o qual nada na comunicação humana escapa a algum tipo de gênero e de que o todo é ponto de partida para dar sentido às partes -, concebe-se que se trabalhe com a leitura crítica do gênero. Isso porque é com base na constatação das formas e peculiaridades do *abstract* que é possível produzi-lo com coerência, conforme já apontado. Para essa atividade, destaca-se o *subcorpus* internacional do *Corpus CISA* e a identificação de características que distinguem os textos. Dentre essas relacionam-se o tipo do *abstract* – informativo, indicativo ou crítico –, o formato – estruturado ou desestruturado –, a extensão média e o modelo IMRD ou a indicação das principais informações acerca da pesquisa. O objetivo pedagógico é a constatação dos “planos de construção textual relativamente fixos” (HOFFMANN, 2015, p. 28) que tipificam o gênero. Também é fundamental que se realce o aspecto de que os *abstracts* apresentam relativa variação, quer a partir das características recém mencionadas para a identificação, quer a partir das regras de submissão dos periódicos. Por isso, é igualmente importante que se trabalhe diretamente com os *sites* dos periódicos internacionais, para os campos específicos em questão; neste caso, Medicina, Nutrição e Farmácia, representados pelo *Annals of Internal Medicine*, pelo *American Journal of Clinical Nutrition* e pelo *British*

Journal of Pharmacology. A finalidade é de promover o reconhecimento das regras de submissão dos periódicos, em relação aos *abstracts* disponíveis para acesso do leitor nos respectivos *sites*. Esse exercício pode, também, fomentar um cotejamento das diferenças entre as publicações especializadas e embasar uma discussão sobre os principais critérios para redação de textos que pretendam publicação em revistas acadêmicas de excelência. Nesse ponto, é oportuno fazer um levantamento breve das origens dos autores, geralmente, relacionadas, na sequência textual, às respectivas identificações. Com isso, deixa-se claro que escrever bem em língua inglesa não tem a ver com a língua cotidianamente falada em nosso país de residência e que, portanto, todos somos capazes de, com dedicação e sistematicidade, desenvolver uma escrita competente em língua inglesa. Também é indispensável proporcionar uma reflexão sobre possíveis alterações à que a síntese de uma pesquisa precisa ser submetida, a fim de se adequar a uma narrativa para um público não especializado, por exemplo. Ao ser levada a cabo com exemplos narrativos, tal tarefa nos alça ao segundo fundamento e ao trabalho específico com nossa unidade de análise.

3. O princípio idiomático de Sinclair (1987) sustenta a continuidade do trabalho pedagógico. Parte-se da noção de que a decisão acerca dos grupos de itens lexicais a adotar, tidos como blocos básicos de construção do discurso, cumpre papel preponderante em termos de definir opções retóricas efetivas. Por isso, para dar sequência ao ensino, exemplificam-se algumas possibilidades de tarefas com o mesmo objetivo – reconhecer e fazer uso dos pacotes lexicais – e com níveis de complexidade variados:
 - (a) Introduzir exemplos de pacotes lexicais a partir do uso do AntConc (ANTHONY, 2014), para que os participantes possam experimentar o trabalho com um *corpus* e entender o quão representativa é a oportunidade de examinar dados autênticos. Além disso, encoraja-se que a prática promova aprendizagem pela descoberta. Para isso, propõe-se orientar os estudantes a alimentar o *software* com o *Corpus* CISA e guiá-los na extração dos pacotes lexicais, com base no passo a passo descrito na subseção 5.4.3 – Do passo a passo analítico - , do capítulo 5. Em

segmento, sugere-se o mesmo exercício com outros *corpora* de textos acadêmicos ou não. Para tanto, indica-se o uso da relação de *corpora* disponíveis online, a partir da tabela ‘*Corpora em Língua Inglesa Disponíveis Online*’, registrada na subseção 4.1.2, do capítulo 4. Com esse exercício, pretende-se conduzir a compreensão do que consistem as colocações estendidas, assim como a constatação de que essas são determinadas pelo uso e de que seu significado é, na maioria das vezes, transparente. Um dos pontos primordiais para a atividade é encorajar a descoberta pela exploração, ou seja, desenvolvê-la de modo que os usuários, com suas palavras, de forma indutiva, expliquem os fragmentos de texto após a extração. Para concluir a tarefa, recomenda-se que o professor faça uso simplificado da síntese de características que tipificam os pacotes lexicais, apresentada na subseção 4.3.2 - Das principais características dos pacotes lexicais - , do capítulo 4;

- (b) Ofertar textos de nosso *subcorpus* internacional com os pacotes lexicais embaralhados (*‘at the end of’* x *‘end at the of’*) para o exercício de reconhecimento;
- (c) Recortar vários fragmentos de textos do *corpus* com lacunas, todos correspondentes ao mesmo pacote lexical no original, para que o feixe de palavras seja identificado, dentre uma sequência de opções que o professor ofereça;
- (d) Apresentar textos do *corpus* com lacunas no lugar de pacotes lexicais variados, para a detecção de possibilidades e comparações posteriores com o original;
- (e) Trabalhar com textos completos, com os pacotes lexicais destacados, porém fora de ordem, a fim de promover a percepção da sequência adequada, para dar sentido ao texto. Com esse exercício, os acadêmicos podem, igualmente, refletir acerca de quais dessas sequências recorrentes de palavras são intercambiáveis e quais não são, bem como sobre os motivos, em termos de coesão e coerência textual, para tais diferenças e

semelhanças;

- (f) Elicitar sinônimos dos feixes lexicais, contendo associações de três a quatro palavras. Esse exercício tem por função trabalhar a compreensão de sentido das colocações estendidas e evidenciar a relevância do item frequência de uso. Para chegar à segunda função, de checar se as expressões elencadas são intercambiáveis ou frequentes na mesma medida, concebemos trabalhar com o *Corpus* CISA, correlacionado ao *software* AntConc (ANTHONY, 2014) ou ao desdobramento pedagógico que apresentaremos na próxima seção do texto, para continuidade da pesquisa;
- (g) Reconhecer as funções retóricas a serem nomeadas após o exercício de identificação dos feixes lexicais. Com isso, pretende-se pontuar *o que aquela sequência de palavras está fazendo ali*. Para o trabalho com o *Corpus* CISA, sugere-se que os objetivos funcionais sejam elencados em um *handout*, para que se relacione a função ao pacote lexical em questão, dado o contexto de uso no texto. Considera-se como mais didático que se trabalhe com um único bloco de objetivos funcionais por lição, para fins de sistematização;
- (h) Desenvolvida a identificação funcional, é a vez de apresentar um quadro com as classificações léxico-gramaticais básicas, para que as colocações estendidas sejam correlacionadas. Para esse exercício, dimensiona-se o uso de um *handout* contendo as classificações dispostas nos gráficos de distribuição léxico-gramatical, da seção 7.1- Do tipo de variabilidade lexical e do comportamento estrutural e funcional dos dados que integram o *corpus* - , do capítulo 7. Seria mais didático, dada a complexidade da tarefa, apresentar os feixes de palavras já em blocos por classificação, para que sejam relacionados às respectivas categorias. Do mesmo modo, pretende-se realçar *como aquela sequência de palavras se organiza*, assim como sistematizar o aspecto de que os feixes lexicais se associam de modos limitados;
- (i) O próximo passo é o trabalho com a porção brasileira do *Corpus* CISA,

para extração dos grupos poli-lexicais e comparação com o segmento internacional dos dados. Para tanto, sugere-se tomar por base, de modo simplificado, os principais aspectos listados na seção 7.2 - Da síntese dos resultados - , do capítulo 7¹⁹⁷:

- A fim de promover reflexão sobre a questão da maior x menor repetição de sequências de palavras e, portanto, sobre uma maior x menor variedade de uso de sequências recorrentes de palavras: apresentar a lista dos dez pacotes lexicais mais frequentes de cada coleção (tabela 18 da seção 7.1 - Do tipo de variabilidade lexical e do comportamento estrutural e funcional dos dados que integram o *corpus* - do capítulo 7), com os respectivos exemplos do contexto de uso a partir das linhas de concordância;
- Para discutir as possíveis implicações de um maior x menor número de realizações de feixes de palavras: expor os alunos a textos de cada *subcorpus* que exemplifiquem a questão;
- Com a finalidade de enfatizar as diferenças retóricas em termos de objetividade de expressão entre os *subcorpora*: disponibilizar *handout* com pacotes lexicais na forma de locuções preposicionadas, desempenhando a função de detalhar aspectos da organização das investigações e desenvolver a tarefa de correlacionar função retórica e composição léxico-gramatical;
- A fim de grifar as distinções entre maior e menor adoção de voz passiva relacionada à promoção de maior agência do objeto do fazer científico: oferecer aos participantes sentenças completas que fazem uso de pacotes lexicais, originalmente, em voz passiva no *subcorpus*, em voz ativa, para o exercício de reescrita, de modo que o agente da ação não seja enfatizado. Problematicar o aspecto das distinções de efeito e discorrer sobre o fato de que a prosa acadêmica é descrita como usuária majoritária de voz passiva.

¹⁹⁷ As conclusões que se seguem, em forma de tarefas pedagógicas, com o objetivo de promover reflexão crítica, são situadas. Não há pretensão de que sejam generalizáveis para o todo da produção científica brasileira. Outrossim, as ponderações feitas são sustentadas por dados autênticos e demonstram que, em relação às porções específicas de língua pesquisadas, as evidências geradas são comprováveis quantitativa e qualitativamente.

Apresentar, por fim, os pacotes lexicais que adotam a construção em dois blocos divididos por acervo;

- Com o propósito de ressaltar o uso de *hedges* ou moduladores do discurso através do emprego da estrutura ‘*anticipatory it pattern*’, como trunfo do acervo brasileiro na condução do foco de interesse do leitor, para a compreensão do estudo: organizar as sentenças das quais as referidas colocações estendidas se originam redigidas de outro modo e propor transformações, para que iniciem com a estrutura em questão;

- A fim de desencadear ponderações sobre as implicações das diferenças nas principais funções dos feixes de palavras nos dois acervos (ferramenta do fazer científico nos textos x justificativa do trabalho acadêmico): distribuir os pacotes lexicais relacionados às principais funções retóricas encontradas em cada *subcorpus*, para que sejam relacionados à função equivalente, com o propósito de realçar as distinções. Conduzir, posteriormente, reflexão acerca de possíveis implicações dessas diferenças.

- (j) Com a finalidade de desenvolver exercícios de escrita para lidar com os aspectos acima, concebe-se a tarefa de relato, em formato de *abstract*, das pesquisas em que os acadêmicos estejam envolvidos, assim como de outras investigações. Para tanto, é importante o uso de um número de pacotes lexicais previamente elencados como preponderantes. Para acadêmicos de outros campos da saúde, recomenda-se que o trabalho seja precedido pelo acesso a periódicos específicos das áreas em questão, com propósito de organização de um *corpus* e extração dos pacotes lexicais mais frequentes. É conveniente organizar processualmente a atividade, com trocas de textos entre pares, revisões e re-escrituras. Desse modo, encoraja-se a interação e o compartilhamento como etapas inerentes à aprendizagem em sala de aula.

É, igualmente, oportuno considerar, também, em sala de aula, 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 vernaculares, os textos internacionais do *Corpus CISA*, por sua vez, são ricos. Por outro lado, é preciso ter cautela para não incorrer em generalizações da conta de um modelo monolítico, já que vários estudos, relatados no capítulo 4, declaram que pacotes lexicais ocorrem e se comportam diferentemente em relação a ambientes disciplinares distintos.

Cumpra 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 acadêmica 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 ao material didático adotado ensinar pelo exercício prático, indutivo e analítico sobre essas distinções, e cabe ao acadêmico aprender a fazer escolhas informadas, a respeito das formas de expressão que dão forma ao seu fazer científico.

Em uma última nota a respeito de ensino, um outro ponto preponderante a considerar é o fato de que, apesar do entusiasmo desta investigação com o trabalho de *corpus*, os acadêmicos aprendem de modos variados. Portanto, que o mesmo não seja percebido como a única opção pedagógica, para pautar o ensino. Com esse respeito, reforçamos o ponto de vista de Biber *et al.* (2004): “que o *corpus* seja nosso aliado, não nosso única opção”. O que é fundamental, de todo modo, é que as tarefas pedagógicas partam de exemplos autênticos, mesmo que não se trabalhe diretamente com *corpora* em sala de aula.

A propósito, uma outra possibilidade de utilizar dados genuínos em um só espaço, portanto com praticidade e didatismo para professor e aluno, é pela adoção de um ambiente virtual de aprendizagem (AVA)¹⁹⁸ *online*, que pode ser abastecido, por exemplo, pelos achados descritos neste estudo. A adoção do componente tecnológico representa um recurso contemporâneo e prático, ao qual se faz referência na próxima seção.

8.2. Das perspectivas futuras e do alcance da pesquisa

¹⁹⁸ Os ambientes virtuais de aprendizagem (AVA) integram as diversas ferramentas para auxiliar a criação e a oferta de cursos mediados por tecnologias, surgidas a partir do desenvolvimento da internet e da popularização do uso do computador (BELMONTI e GROSSI, 2010).

Ao chegarmos ao final da pesquisa, é pertinente pensar acerca de dois aspectos: que desdobramentos futuros terão lugar a partir da investigação e o que poderia ter sido organizado diferentemente. Iniciemos pelas projeções.

Um desdobramento futuro concebido é a criação de um AVA. O propósito é dar suporte à redação de *abstracts* no segmento da saúde e, assim, apoiar 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, igualmente, a promover a sistematização didática das conclusões da pesquisa, bem como, a contribuir com os avanços do campo de ensino e aprendizagem de EAP.

Tal AVA comportaria, futuramente, uma ferramenta de suporte à escrita, para apoiar a construção de diferentes tipos de *abstracts*, voltados para artigos, monografias de conclusão de cursos de graduação, dissertações, teses e trabalhos para apresentações em eventos do campo da saúde. Caberia ao usuário eleger o tipo e a subárea para seu *abstract*. Feitas as escolhas, o autor acessaria um *template* e, ao concluir seu texto, receberia um *feedback* da própria ferramenta, sem, no entanto, alterações automáticas no texto original, para possibilitar a verificação e comparação das sugestões feitas.

O texto ficaria armazenado no sistema e alimentaria a base de *feedbacks*, formando um *corpus*. O SCIPO¹⁹⁹ (FELTRIM *et al.*, 2003), por exemplo, também é um portal de escrita acadêmica para língua inglesa disponível aos usuários brasileiros. A ferramenta, por sua vez, ocuparia 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*, a exemplo do trabalho de Torres (2015). O estudo dessa pesquisadora “propôs a criação de recursos e ferramentas de suporte à escrita no nível lexical

¹⁹⁹ <http://www.nilc.icmc.usp.br/~scipo/> (Último acesso em 20/03/16.)

como primeiro passo para a melhoria da qualidade linguística” (TORRES, *op. cit.*, p. v), no caso, em relação a textos produzidos em língua portuguesa por falantes de espanhol como primeira língua, como é nossa proposta executar para o benefício de usuários brasileiros que necessitem produzir textos em língua inglesa.

Por ora, relacionam-se os aspectos idealizados como principais componentes do ambiente:

- Área de apresentação da ferramenta, seus objetivos e público-alvo;

- Tutorial para usuários,

- Exemplificação de *abstracts* típicos da área da saúde a partir de aspectos mais genéricos (modelos de Swales, 1990 e Swales e Feak, 2009) e das três revistas especializadas escolhidas para o desenvolvimento deste estudo;

- Apresentação sucinta do que constituem pacotes lexicais, além de exemplos a partir dos principais *lexical bundles* aferidos no estudo, bem como a descrição de aspectos relevantes dos dez mais frequentes;

- Inclusão de espaço para um *corpus* embutido;

- Área de cadastro de arquivos de *abstracts*;

- Preservação do texto original do usuário e, a partir dele, produção da revisão, apontando, em cores, o que pode ser alterado para melhorá-lo;

- Espaço para avaliação da ferramenta e do AVA pelos usuários, bem como um sistema de registro de suas contribuições;

- Sistema de correio eletrônico conjugado, informando aos administradores sobre a ocorrência de mensagens e comentários;

- *Links a sites* com tutoriais para escrita acadêmica, a fim de encorajar o desenvolvimento da aprendizagem e a autoria dos usuários nos processos de escrita;
- Espera para vídeos de aulas e para roteiros de apresentações em congressos.

Vale acrescentar que o AVA representaria um recurso a ser conjugado aos esforços pedagógicos registrados anteriormente, não um substituto. Uma das proposições que norteiam a concepção do ambiente virtual é alavancar a autonomia dos usuários. Outrossim, a capacidade de redigir um texto e de fazer escolhas informadas, acerca dos itens lexicais a serem adotados para a redação acadêmica, é um legado do usuário. Portanto, parodiando Biber, que a tecnologia seja nossa aliada, não nossa única opção.

Por fim, chegamos às considerações sobre o que poderia ter sido organizado diferentemente. Quando se recorta o formato de uma pesquisa, fatalmente, incorrem-se em escolhas que deixam de fora outras possibilidades. Neste nosso caso, optou-se por investigar apenas as sub-áreas biomédicas de Medicina, Nutrição e Farmácia. Certamente, é importante que se realizem pesquisas abrangendo outras sub-áreas, a fim de ampliar a representatividade dos achados para outros campos das Ciências da Saúde.

Outro ponto a considerar é que o trabalho foi executado com um *corpus* de pequenas dimensões, embora o acervo tenha se provado efetivo para dar conta das questões postuladas para o estudo e tenha gerado evidências valiosas a respeito dos *subcorpora* delimitados. A questão do tamanho ideal de um *corpus*, em Linguística de *Corpus*, é sempre um assunto polêmico. Entretanto, acredita-se que a representividade e a escolha criteriosa de fontes e de textos pode ser uma vantagem importante. Mesmo assim, é sempre valioso analisar mais dados léxico-gramaticais em se tratando de uma investigação em LC. A tarefa de compilação e organização de um *corpus* de grandes proporções, por sua vez, representa uma empreitada árdua para um pesquisador que trabalha sozinho. Quase todos os estudos a que referimos ao longo deste trabalho foram obra de equipes com vários participantes. É objetivo, portanto, constituir uma pequena equipe de pesquisa para dar continuidade à pesquisa, tanto no sentido de ampliar o *corpus*, quanto, a fim de desenvolver o desenho do AVA e avaliar suas melhores opções de implementação.

Um terceiro aspecto relacionado ao escopo do trabalho é a questão de que, para estabelecer a extensão das ocorrências que seriam extraídas do *corpus*, adotou-se o critério de quatro palavras gráficas. Para delimitar a linha de corte, seguiu-se o princípio norteador de que os pacotes lexicais de três palavras são muito frequentes e geralmente se abrigam em

sequências de quatro palavras. No entanto, mesmo que os pacotes lexicais contendo mais de quatro palavras sejam mais raros, idealiza-se identificá-los futuramente. Isso porque é didaticamente oportuno que também se apresentem sequências mais longas, comparadas às mais curtas, para auxiliar os aprendizes no reconhecimento das porções lexicais que compõem os textos acadêmicos.

Para finalizar, a fim de investigar a composição dos *abstracts* em Ciências da Saúde, em língua inglesa, este estudo adotou uma unidade analítica fragmentária, ao eleger os pacotes lexicais como forma de aproximação do *corpus* e, com isso, em grande parte, subliminal aos olhos do acadêmico. No entanto, como demonstrado, é possível analisar aspectos significativos sobre a identidade discursiva dos textos acadêmicos com base no uso dessas associações de porções lexicais, fragmentos que tenderiam a passar despercebidos se não tivéssemos uma perspectiva extensiva de observação de usos da língua. O relato dos estudos científicos, na verdade, é capaz de fornecer indícios não apenas sobre as investigações no que diz respeito a seu caráter experimental, mas, ainda, sobre como produzimos ciência, sobre como podemos ser percebidos na comunidade internacional e, com isso, sobre o quanto pertencemos ou não a essa comunidade. Portanto, o esforço do trabalho foi no sentido de tornar explícitos esses fragmentos, relativamente *invisíveis*, e de apontar seu caráter primordial para a produção fluente da escrita. Como se pode perceber, o texto especializado também é especificado por elementos convencionais e discursivos que vão além das suas terminologia mais típicas. Ao concluir a tese, espera-se ter contribuído para que outros pesquisadores e educadores linguísticos possam contar com indicativos autênticos sobre o tema da produção de *abstracts* nos domínios das Ciências da Saúde. É expectativa que se vislumbrem, igualmente, insumos para novas pesquisas. Ao mesmo tempo, estima-se auxiliar na promoção da qualidade do ensino e que esse tenha o caráter de dar sentido ao todo pelo fragmentário, para que, pela língua inglesa, seja possível agir no mundo da ciência de forma plena.

9. REFERÊNCIAS BIBLIOGRÁFICAS

AARTS, J. Does corpus linguistics exist? Some old and new issues. In: BREIVIK, L. E. ; HASSELGREN, A. [eds.] **From the COLT's Mouth ... and Others': Language Corpora Studies in Honour of Anna-Brita Stenström**. Amsterdam: Rodopi, p. 1–19, 2002.

ABNT. **Normas Técnicas 2015**. Disponível em: <<http://www.abnt.org.br/normalizacao/lista-de-publicacoes/abnt/category/142-novembro>>. Acesso em: 26 nov. 2015.

AD HOC GROUP FOR CRITICAL APPRAISAL OF MEDICAL LITERATURE. A Proposal for More Informative Abstracts of Clinical Articles. **Annals of Internal Medicine**, v. 106, p. 598-604, 1987.

ADOLPHS, S. ; BROWN, B. ; CARTER, R. ; CRAWFORD, P. SAHOTA, O. Applied clinical linguistics: corpus linguistics in health care settings. **Journal of Applied Linguistics**, v. 1, p. 9–28, 2004.

ALTENBERG, B. 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, p.101-122, 1998.

ALTENBERG, B. Recurrent verb-complement constructions in the London-Lund Corpus in: AARTS, J.; HAAN, P. ; OOSTDIJK, N. [eds.] **English language corpora: design, analysis and exploitations**. Rodopi: Amsterdam, p. 227-245, 1993.

ALTENBERG B. [org.] **English corpus linguistics: Studies in honour of Jan Svartvik**. London: Longman, 1991.

ALUÍSIO, S.; ALMEIDA, G. M. B. O que é e como se constrói um corpus? Lições aprendidas na compilação de vários corpora para a pesquisa linguística. In: **Calidoscópico**, v. 4 (3), setembro/dezembro, 2006. Disponível em: <<http://revistas.unisinos.br/index.php/calidoscopio/article/view/6002>>. Acesso em: 28 nov. 2015.

ALUÍSIO, S.M.; GANTENBEIN, R.E. Educational Tools for Writing Scientific Papers. In: **VIII Simpósio Brasileiro de Informática na Educação**, ITA, p. 239-253, 1997.

ALUÍSIO, S.M; OLIVEIRA, Jr., O. A Case-Based Approach for Developing Writing Tools Aimed at Non-native English Users. **Lecture Notes in Artificial Intelligence 1010**, p. 121-132, 1995.

ANSI Z39.14. **The American National Standard for Writing Abstracts**. New York: American National Standards Institute, 1979. Disponível em: http://www.niso.org/apps/group_public/project/details.php?project_id=124. Acesso em: 20 out. 2015.

ANTHONY, L. **AntConc (Version 3.4.1) [Computer Software]**. Tokyo, Japan: Waseda University, 2014. Disponível em: <<http://www.antlab.sci.waseda.ac.jp/>>. Acesso em: 28 dez. 2015.

ANTHONY, L. A critical look at software tools in *corpus* linguistics. **Linguistic Research**, v. 30(2), p. 141-161, 2013.

ANTHONY, L. Developing a Freeware, Multiplatform Corpus Analysis Toolkit for the Technical Writing Classroom. **IEEE Transactions on Professional Communication**, v. 49 (3), p. 275-286, 2006.

ANTHONY, L.; CHUJO, K. ; OGHIGIAN, K. A novel, web-based, parallel concordancer for use in the ESL/EFL classroom. In: NEWMAN, J.; BAAYEN, H. ; RICE, S. [eds.] **Corpus-based Studies in Language Use, Language Learning, and Language Documentation**. NY: Rodopi, p. 123-138, 2011.

ARAÚJO, J. C.; SOUSA, C. T. [orgs.] **Gêneros textuais e comunidades discursivas: um diálogo com John Swales**. Belo Horizonte: Editora Autêntica, p. 33-46, 2009.

ARAÚJO, A. D. Uma análise d eorganização discursiva de resumos na área de Educação. **Revista do GELNE – Grupo de Estudos Linguísticos do Nordeste**, v. 1, p. 26-30, 1999.

ASTON, G. Small and large corpora in language learning. **Paper presented at the PALC Conference, University of Lodz, Poland**, April 1997.

ATHANASAKOU, V. ; EL-HAJ, M. ; RAYSON, P. ; WALKER, M.; YOUNG, S. Computer-based Analysis of the Strategic Content of UK Annual Report Narratives. **American Accounting Association Annual Meeting**, August 2-6, 2014, Atlanta, USA, 2014.

ATKINS, S.; CLEAR, J.; OSTLER, N. Corpus design criteria. **Literary and Linguistic Computing**, v.7, p. 1-16, 1992.

BAKER, P. Unnatural acts' Discourses of homosexuality within the House of Lords debates on gay male law reform. **Sociolinguistics**, v. 8 (1), p. 88-106, 2004.

BAKER, P. ; GABRIELATOS , C. ; MCENERY, A. **Discourse Analysis and Media Attitudes: The representation of Islam in the British Press**. Cambridge: Cambridge University Press, 2013.

BAKHTIN M. O problema dos gêneros discursivos. In: **Estética da criação verbal**. Tradução Paulo Bezerra . 4a Ed. São Paulo: Martins Fontes, 2003.

BAKHTIN M. **Marxismo e filosofia da linguagem**. São Paulo: Hucitec, 1999.

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

BAKHTIN, M. **Speech Genres and Other Late Essays**. Trans. by Vern W. McGee. Austin/Texas: University of Texas Press, 1986.

BARROS, L. A. Aspectos Epistemológicos e Perspectivas Científicas da Terminologia. **Ciência & Cultura**. v. 58 (2) , p. 22-26. Abr./Jun., 2006.

BARROS, L. A. **Curso Básico de Terminologia**. São Paulo: Ed. da Universidade de São Paulo, 2004.

BARROS, D. **Teoria Semiótica do Texto**. São Paulo: Ática, 1990.

- BARTHES, R. **O prazer do texto**. Tradução J. Guinsburg. 4. ed. São Paulo: Perspectiva, 1973/1996.
- BAZERMAN, C. **Shaping Written Language**. The University of Wisconsin Press: Madison/Wisconsin, 1988.
- BEAUFORT, A. **Transferring Writing Knowledge to the Workplace: Are We on Track? *Expanding Literacies: English Teaching and the New Workplace***. [eds.] GARAY, M. S. Garay ; BERNHARDT, A. Bernhardt. Albany: SUNY, p. 179-99, 1998.
- BEAUGRANDE, R. A. **New Foundations for a Science of Text and Discourse**. Greenwich, CT: Ablex, 1997.
- BEAUGRANDE, R. A.; DRESSLER, W. U. **Einführung in die Textlinguistik**. Tübingen: Niemeyer, 1981.
- BEDNAREK, M. **Evaluation in Media Discourse: Analysis of a Newspaper Corpus**. London/New York: Continuum, 2006.
- BELMONTI, V.; GROSSI, M. G. R. **Ambientes virtuais de aprendizagem: um panorama da produção nacional**. Disponível em: www.abed.org.br/congresso2010, Belo Horizonte, 2010. Acesso em: janeiro de 2014.
- BERBER SARDINHA, A. P. **Linguística de *Corpus***. São Paulo: Manole, 2004.
- BERBER SARDINHA, A. P. Análise de Gêneros e Linguística de *Corpus*: Identificação das unidades internas do gênero por meio de padronização lexical. **DIRECT Papers**, v. 51, p. 1-30, 2003. Disponível em: <http://www2.lael.pucsp.br/direct/DirectPapers51.pdf> Acesso em: 20 nov. 2015.
- BERBER SARDINHA, A. P. O corpus de aprendiz Br-I-CLE. **Intercâmbio**, v. 10, p. 227-239, 2001.
- BERBER SARDINHA, A. P. Linguística de Corpus: Histórico e Problemática. **D.E.L.T.A.**, v.16 (2), p. 323-367, 2000. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502000000200005 Acesso em: 20 nov. 2015.
- BEVILACQUA, C. R. **Unidades fraseológicas Especializadas: descripción y regras de formación em el ámbito de la energía solar**. Tese de Doutorado. Barcelona: Instituto Universitário de Linguística Aplicada (IULA), UPF, 2004.
- BHATIA, V. K. **Analysing Genre: Language Use in Professional Settings**. London: Longman, 1993.
- BHATIA, V. ; LANGTON, N. ; LUNG, J. Legal discourse: Opportunities and threats for corpus linguistics. In: CONNOR, U. ; UPTON, T. [eds.] **Discourse in the Professions: Perspectives from Corpus Linguistics**. Amsterdam/Philadelphia: John Benjamins, p. 203–231, 2004.
- BIASI-RODRIGUES, B.; ARAÚJO, J. C.; SOUSA, S. C. T. [org.] **Gêneros textuais e comunidades discursivas: um diálogo com John Swales**. Belo Horizonte: Autêntica, 2009.

BIBER, D. **University Language – A corpus-based study of spoken and written registers.** Amsterdam/Philadelphia: John Benjamins Publishing Company, 2006.

BIBER, D. Investigating language use through corpus-based analyses of association patterns. In: BARLOW, M.; KEMMER, S. [eds.] **Usage based models of language.** Stanford: CSLI Publications, p. 287-314, 2000 [Reprint of 1996 article in *International Journal of Corpus Linguistics*.]

BIBER, D. **Dimensions of Register Variation – A Cross-Linguistic Comparison.** Cambridge: Cambridge University Press, 1995.

BIBER, D. Methodological issues regarding corpus-based analyses of linguistic variation. **Literary and Linguistic Computing**, v.5, p.257-269. 1990.

BIBER, D. **Variations across speech and writing.** Cambridge University Press: Cambridge, 1988.

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

BIBER, D. ; CONRAD, S. **Register, genre, and style.** Cambridge: Cambridge University Press, 2009.

BIBER, D. ; CONRAD, S. Lexical bundles in conversation and academic prose. In: HASSELGARD, H. ; OKSEFJELL, S. [orgs.] **Out of Corpora: Studies in Honour of Stig Johansson.** Amsterdam/Atlanta, GA: Rodopi, p. 181-190, 1999.

BIBER, D.; CONRAD, S.; CORTES, V. *If you look at ...*: Lexical bundles in university teaching and textbooks. **Applied Linguistics**, v. 25 (3), p. 371–405, 2004.

BIBER, D., CONRAD, S.; CORTES, V. Lexical bundles in speech *and writing: An initial taxonomy.* In WILSON, A. ; RAYSON, P.; MCENERY, T. [eds.] **Corpus linguistics by the lune: A festschrift for Geoffrey Leech.** Frankfurt/Main: Peter Lang, p.71-93, 2003.

BIBER, D.; CONRAD, S. ; REPPEN, R. **Corpus linguistics: investigating language structure and use.** Cambridge: Cambridge University Press, [1998] 2000.

BIBER, D.; FINEGAN, E. An Initial Typology of English Text Types. In: AARTS, J; MEIJS, W. [eds.] **Corpus Linguistics II: New Studies in the Analysis and Exploitation of Computer Corpora.** Amsterdam: Rodopi, p. 19-46, 1986.

BIBER, D. ; GRAY, B. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. **Journal of English for Academic Purposes**, v. 9 (1), p. 2-20, 2014.

BIBER, D.; GRAY, B. **Discourse Characteristics of Writing and Speaking Test Types on the TOEFL iBT Test: A Lexico-Grammatical Analysis.** Princeton: ETS, 2013^a.

BIBER, D. ; GRAY, B. **Nominalizing the verb phrase in academic science writing.** ACLW: Cambridge University Press, 2013^b. Chapter 5.

BIBER, D. ; GRAY, B. Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. **Journal of English for Academic Purposes**, v. 9, p. 2-20, 2010.

BIBER, D. JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. **Longman**

grammar of spoken and written English. Harlow: Pearson, 1999.

BLAIS, E. Le phraseologisme. Une hypothèse de travail. **Terminologies Nouvelles**. Bélgica: RINT, n. 10, p. 50-56, 1993.

BLÜHDORN, H.; ANDRADE, M. L. Tendência recentes da linguística textual na Alemanha e no Brasil. **Revista de Filologia e Linguística Portuguesa**, v 7, p. 13-48. São Paulo: Humanitas/ FFLCH-USP, 2005.

BOLINGER, D. Meaning and memory. **Forum Linguisticum**, v.1, p. 1-14, 1976.

BORDIEU, P. **Language and symbolic power**. Oxford: Polity Press.1991.

BOURIGAULT, D.; SLODZIAN, M. Pour une terminologie textuelle. **Terminologies Nouvelles Terminologie et Intelligence Artificielle**. n. 19, p. 29-32, 1999.

BREZINA, V. **Statistics in Corpus linguistics**: Web resource, available from <http://corpora.lancs.ac.uk/stats>, 2014. Acessado em 01 jan. 2016.

BRINKER, K. **Linguistische Textanalyse: eine Einführung in Grundbegriffe und Methoden**, 6. Überar und erw. Aufl. Berlin: Erich Schmidt, 2005.

BRINKER, K. **Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden**. Berlin, Grundlagen der Germanistik, 29, 1985.

BUHLMANN, R.; FEARN, A. **Handbuch des Fachsprachenunterrichts**. Langenscheidt, 1987.

BYRD, P. ; COXHEAD, A. On the other hand: Lexical bundles in academic writing and in the teaching of EAP. **University of Sydney Papers in TESOL**, v. 5, p. 31-64, 2010.

CABRÉ, M. T. Textos especializados y unidades de conocimiento: metodología y tipologización. In: GARCIA PALACIOS, J.; FUENTES, M. T. [eds.] **Texto, Terminología y traducción**. Salamanca: Almar, p. 15-36, 2002.

CABRÉ, M. T. El discurso especializado o la variación funcional determinada per la temática: nuevas perspectivas. In: CABRÉ, M. T. **La terminología: representación y comunicación; elementos para una teoría de base comunicativa y otros artículos**. Barcelona: IULA, p. 151-173, 1999.

CASELI, H. M.; RAMISCH, C.; GRAÇAS VOLPE NUNES , M. ; VILLAVICENCIO, A. Alignment-based extraction of multiword expressions. **Language Resources and Evaluation**, p. 1-20, 2009.

CHAN, S. K; FOO, S. Interdisciplinary Perspectives on Abstracts for Information Retrieval. **Iberica**, v. 8, p. 101-124, 2004.

CHEN, Y. ; BAKER, P. Lexical Bundles un L1 and L2 Academic Writing. **Language Learning & Technology**, v. 14 (2), June, p. 30–49, 2010.

CHOMSKY, N. **Knowledge of language: Its nature, origin, and use**. Cambridge: MIT Press, 1986.

CIAPUSCIO, G. **Textos especializados y terminología**. Barcelona: IULA, 2003.

CIAPUSCIO, G. La terminología desde el punto de vista textual: selección, tratamiento y variación. **Organon**, Porto Alegre, v. 12 (26), p. 43-65, 1998.

CIAPUSCIO, G. **Tipos Textuales**. Buenos Aires: Eudeba, 1994.

CIAPUSCIO, G.; KUGUEL, I. Hacia una tipología del discurso especializado: aspectos teóricos y aplicados. In: GARCÍA PALACIOS, J. ; FUENTES, M. T. [eds.] **Entre la terminología, el texto y la traducción**. Salamanca: Almar, p. 37-73, 2002.

CILVETTI, L. D.; PÉREZ, I. K. L. Textual and language flaws: Problems for Spanish doctors in producing abstracts in English. **Ibérica**, v.11, p. 61-79, 2006.

CONDAMINES, A. Linguistique de corpus et terminologie. In: **Langages**, 157, p. 36-47, 2005. Disponível em: <http://w3.univtlse2.fr:8880/erss/index.jsp?perso=acondami&subURL=Langagesdef.pdf>. Acesso em: 15 Jan. 2015.

CONNOR, U.; PRECHT, K.; UPTON, T. **Business English: Learner data from Belgium, Finland and the U.S. In Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching**. In: GRANGER, S. ; HUNG, J. ; PETCH-TYSON, S. [eds.] Amsterdam: John Benjamins, p. 175-194, 2002.

CONRAD, S.; BIBER, D. The Frequency and Use of Lexical Bundles in Conversation and Academic Prose. In: TEUBERT, W. ; MAHLBERG, M. [eds.] **The corpus approach to lexicography, Thematischer Teil von Lexicographica**. Internationales Jahrbuch für Lexicographie, 20, 2004, 56-71, 2005.

CONTE, M. E. **La lingüística testuale**. Milano: Feltrinelli Económica, 1977.

CORTES, V. A comparative analysis of lexical bundles in academic History writing in English and Spanish. **Corpora**, v. 1(1), p. 43-57, 2008.

CORTES, V. Teaching lexical bundles in the disciplines: An example from a writing intensive history class. **Linguistics and Education**. v. 17, p. 391-403, 2006.

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

CORTES, V. Lexical bundles in Freshman composition. In: REPPEN, R. ; FITZMAURICE, S. M.; BIBER, D. [eds.] **Using corpora to explore linguistic variation**. Amsterdam and Philadelphia: John Benjamins, p. 131-146, 2002.

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

COWIE, A. P. [org.] **Phraseology – Theory, Analysis, and Applications**. Oxford: Clarendon, 1998.

CREMMINS, E. T. **The Art of Abstracting**. Philadelphia: ISI Press, 1982.

CRYSTAL, D. **English as a global language**. Cambridge: Cambridge University Press, 2003.

DAMIN, C. P. ; ALVES, I. M. R. A. ; BECKER, J. L. ; SOUZA, J. A. C. ; GIERING, M. A.; ALBÉ, M. A. ; KISSAMANN, S. & MELLO, V. H. D. **Leitura e produção de textos de comunicação da ciência**. São Leopoldo: Editora UNISINOS, 2013.

DAVOODIFARD, M. Functions and hedges in English and Persian academic discourse: effects of culture and the scientific discipline. **ESP Across Cultures**, n. 5, p. 23-48, 2008.

DAYRELL, C. Padrões léxico-gramaticais na especificação de propósito e resultado em abstracts de artigos científicos: aplicações no ensino de EAP e na construção de ferramentas de suporte à escrita científica . In: IBANOS, A. M. T. *ET AL*. [orgs.] **Pesquisas e perspectivas em linguística de corpus**. Campinas: Mercado de Letras, p. 303-346, 2015.

DAYRELL, C. Anticipatory it in English abstracts: a corpus-based study of non-native student and published writing. In: GOZDZ-ROSKOWSKI S. [org.] **Studies in Languages: Explorations across Languages and Corpora**. Łódź Studies in Language. 1ed. Frankfurt am Main: Peter Lang, , v. 24, p. 581-598, 2011.

DAYRELL, C. Corpora no Ensino do Inglês Acadêmico: Padrões Léxico-Gramaticais em Abstracts de pós-graduandos brasileiros. In: VIANA, V.; TAGNIN, S. E. O. [orgs.] **Corpora no ensino de línguas estrangeiras**. São Paulo: Hub Editorial Ltda., p. 137-171, 2010.

DAYRELL, C. 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^a**. Disponível em: <<http://www.liv.ac.uk/english/CL2009/index.htm>>. Acesso em: 04 jul. 2015.

DAYRELL, C. “Anticipatory ‘it’ in English abstracts: A corpus-based study of non-native student and published writing”. Paper presented at the **7th International Conference on Practical Applications in Language and Computers (PALC 2009)**, Lodz University Conference Centre in Lodz, Poland, 6th to 8th April, 2009.

DAYRELL, C . Sense-related verbs in English scientific abstracts: a corpus-based study of students’ writing. **ESP Across Cultures** , n. 6, p. 61-78, 2008.

DAYRELL, C ; ALUISIO, S. Using a comparable corpus to investigate lexical patterning in English abstracts written by non-native speakers. In: LREC 2008, 2008, Marrakech. **Proceedings of the LREC 2008 Workshop Building and Using Comparable Corpora**, 2008.

DAYRELL, C.; URRY, J. Mediating climate politics: The surprising case of Brazil. **European Journal of Social Theory**, v. 18 (3) p. 257-273, April 22, 2015.

DE COCK, S. Repetitive phrasal chunkiness and advanced EFL speech and writing. In: MAIR, C. ; HUNDT, M. [eds.] **Corpus Linguistics and Linguistic Theory**. Amsterdam: Rodopi, p. 51–68, 2000.

DE COCK, S. ; GRANGER, S. ; LEECH, G.; MCENERY, T. An automated approach to the phrasicon on EFL Learners. In: GRANGER, S. [ed.] **Learner English on Computer**. London/New York: Addison Wesley, Chapter 5, 1998.

DIKI-KIDIRI, M. La terminología cultural: fundamento de una verdadera localización.VIII **SIMPOSIO IBEROAMERICANO DE TERMINOLOGÍA: LATERMINOLOGÍA, ENTRE LA GLOBALIZACIÓN Y LA LOCALIZACIÓN**, Cartagena, Colômbia, 28-31

de outubro de 2002. Disponível em: <<http://www.ritem.net/revista>>. Acesso em: 04 set. 2015.

DUDLEY-EVANS, T. English for Specific Purposes. In: CARTES, R. ; NUNAN, D. [eds.] **The Cambridge guide to teaching English to speakers of other languages**. Cambridge: Cambridge University Press, P. 131-136, 2001.

DUDLEY-EVANS, T. ; ST. JOHN, M. J. **Developments in English for Specific Purposes - A multi-disciplinary approach**. Cambridge: Cambridge University Press, 1998.

DURANTI, A. **Linguistic anthropology**. Cambridge University press: Cambridge/UK, 1997.

DUTRA, D. P. ; BERBER SARDINHA, T. Pacotes Lexicais em *Corpora* de Aprendizizes. In: IBAÑOS, A. M. ; MOTIN, L. P. ; SARMENTO, S. ; BERBER SARDINHA, T. [orgs.] **Pesquisas e Perspectivas em Linguística de Corpus**. Porto Alegre: Mercado de letras, p. 57-80, 2015.

ELLIS, R. **Second language acquisition**. OUP: Oxford, 1997.

ENDRES-NIGGEMEYER B. A procedural model of an abstractor at work. In: **International Forum of Information and Documentation**, v. 15 (4,) p. 3-15, 1990.

ENDRES-NIGGEMEYER B. Referierregeln und Referate- Abstracting als regelsgesteuerter Textverarbeitungsprozess. in **Nachrichten für Dokumentaristen**, v. 36 (1), p. 38-50, 1985.

ENDRES-NIGGEMEYER B.; WAUMANS W.; YAMASHITA. Protocol analysis of non-native abstractors. **Text** , v. 11 (4), p.523-552, 1991.

FADANELLI, S. B.; FINATTO, M. J. A Arquitetura de um glossário terminológico inglês-português na área da Eletrotécnica. **Linguamática**, v.7 (1), p. 67-71, julho de 2015.

FAIRCLOUGH, N. **Language and Power**. Harlow: Longman Group UK Limited, 1989.

FAULSTICH, E. *A Socioterminologia na comunicação científica e técnica*. **Revista Ciência e Cultura**, São Paulo, v. 58 (2), p. 27-31, 2006.

FAULSTICH, E. Socioterminologia: mais que um método de pesquisa, uma disciplina. **Ciência da Informação**, v. 24 (3), p. 1-14, 1995.

FÁVERO, L. L.; KOCH, I. V. **Linguística textual: uma introdução**. São Paulo: Cortez, 1983.

FELBER, H. ; PITCH, H. **Métodos de Terminografía y principios de investigación terminológica**. Madrid: C.S.I.C. , Instituto Miguel de Cervantes. 1984.

FELTRIM, V. D.; PELIZZONI, J. M.; TEUFEL, S.; NUNES, M.G.V.; ALUÍSIO, S. M. Applying Argumentative Zoning in an Automatic Critiquer of Academic Writing. In: **Proceedings of the 17th Brazilian Symposium on Artificial Intelligence**, São Luis-MA, Brazil. *Lecture Notes in Artificial Intelligence*, 3171, Springer, p. 214-223, 2004.

FELTRIM, V. D.; ANTIQUEIRA, L. ; NUNES, M. das G. V. . **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.

FENG, H. *A corpus-based study of research grant proposal abstracts*. **Perspectives: Working Papers in English and Communication**, v.17 (1), p.1-24, Spring 2006.

FILLMORE, C. 'Corpus linguistics' or 'computer corpus linguistics'. In: SVARTVIK, J. [org.] *Directions in Corpus Linguistics*. **Proceedings of Nobel Symposium 82**, Stockholm, 4-8 August 1991. Berlin, New York: De Gruyter, 1992.

FINATTO, M. J. B. Dos dicionários aos modos de perceber a língua: contribuições de um trabalho pioneiro. In: TEIXEIRA, M. ; FLORES, V. N. [orgs] **O sentido na linguagem: uma homenagem à professora Leci Borges Barbisan**. (1ed.) Porto Alegre - RS: EDIPUCRS, v. 1 , p. 17-32, 2012.

FINATTO, M. J. B.. 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, v. 01, p. 153-172, 2011^a.

FINATTO, M. J. B. Complexidade Textual Em Artigos Científicos: Contribuições para o Estudo do Texto Científico em Português. **Organon (UFRGS)**, v. 50, p. 30-45, 2011^b.

FINATTO, M. J. B. [org.] **Linguagens especializadas em corpora: modos de dizer e interfaces**. 1ed.Porto Alegre - RS: EDIPUCRS, v. 01, p. 152-182, 2010.

FINATTO, M. J. B. Exploração terminological com apoio informatizado: perspectivas, desafios e limites. In: ISQUERDO, A. N. ; ALVES, I. M. **As Ciências do Léxico – Lexicologia, Lexicografia, Terminologia**. Campo Grande: Humanitas, v. 03, p. 447-58, 2007.

FINATTO, M. J. B. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, A. N. ; KRIEGER, M. G. [orgs.] **As ciências do léxico: lexicologia, lexicografia, terminologia, volume II**. Campo Grande: Ed. UFMS/ PPG-Letras UFRGS, v. 01, p. 353-363, 2004.

FINATTO, M. J. B. . Conhecendo um software tipo parser: instrumento auxiliar para a pesquisa terminológica e sua utilização pelo estudante de Letras tendo em vista a elaboração de glossários especializados. In: LIMA, Marília Santos; CHITTONI, P, R. [org.] **Canadá e Brasil: questões de ensino e terminologia**. Porto Alegre: UFRGS, Núcleo de Estudos Canadenses, ABECAN, v. 1, p. 55-70, 2001.

FINATTO, M. J. B.; EVERS, A.; ALLE, C. M. O.; ALENCAR, M.C. Das terminologias às construções recorrentes: um percurso de estudos sobre linguagens especializadas. **Revista Íkala Revista de Lenguaje y Cultura**, v. 15, p. 223-258, 2010.

FINATTO, M. J. B. ; EVERS, A. ; ALLE, C. M. O. Para além das terminologias: estudos de convencionalidade em linguagens científicas. In: PERNA, C.L. B. ; DELGADO, H.O. K.; FINATTO, M. J. B. **Linguagens Especializadas em Corpora**. Porto Alegre: EDIPUCRS, p. 152-182, 2010.

FINATTO, M. J. B. ; ISQUERDO, A. N. [orgs.] **As Ciências do Léxico Lexicologia Lexicografia Terminologia volume IV**. 1. ed. Campo Grande - MS: Editora de UFMS, 2008.

FINATTO, M. J. B. ; ZILIO, L. [orgs.] **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, 2015. Disponível em: <<http://www.ufrgs.br/ppglettras/>>. Acesso em: 04 Jun. de 2015.

FIRTH, J. R. **Papers in Linguistics 1934-1951**. London: Oxford University Press. 1957.

FISH, S. Interpreting the Variorum. *Critical Inquiry*, v. 2 (3), p. 465-485, 1976.

FLOWERDEW, J. Introduction: Approaches to the Analysis of Academic Discourse in English In: FLOWERDEW, J. [ed.] **Academic Discourse**. New York: Routledge, p.1-17, 2002.

FLOWERDEW, J. ; FOREST, R. In: CHARLES, M. ; PECORARI, D. ; HUNSTON, S. [eds.] **Academic writing: At the interface of corpus and discourse**. London: Continuum. p.15-36, 2009.

FLOWERDEW, J.; MILLER, L. Dimensions of academic listening. In: CELCE-MURCIA, M. ; BRINTON, D. M. ; SNOW, M. A. Snow [eds.] **Teaching English as a second or foreign language** (4th ed.) Boston, MA: National Geographic Learning/Cengage Learning, p.90-103, 2014.

FLOWERDEW, L. **Corpus-based Analyses of the Problem-Solution Pattern: A Phraseological Analysis**. Amsterdam/Philadelphia: John Benjamins, 2008.

FLOWERDEW, L. 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, p. 11-36, 2004.

FRANCIS, W. N. Language *Corpora* BC. In: SVARTVIK, J. [ed.] Directions in Corpus Linguistics. **Proceedings of Nobel Symposium 82, Stockholm**. Berlin/ New York, p. 17-32, 1992.

FRANCIS, W. N. ; KUCERA, H. **Frequency analysis of English usage: Lexicon and grammar**. Boston: Houghton Mifflin, 1982.

FREIRE, P. **A importância do ato de ler: em três artigos que se completam**. São Paulo: Autores Associados, 1989.

FREITAS, A. L. P. ; MACHADO, Z. F. Noções Fundamentais: A Organização Da Tomada De Turnos Na Fala-Em-Interação. In: LODER, L. L.; JUNG, N. M. [orgs.] **Fala-em-interação social: uma introdução à Análise da Conversa Etnometodológica**. 1a ed. Porto Alegre: Mercado das Letras, v. 01, p. 59-94, 2008.

GENOVÊS Jr., L.; FELTRIM, V.D.; DAYRELL, C. ; ALUÍSIO, S. Automatically detecting schematic structure components of English abstracts. In Proceedings of the RANLP 2007, **Workshop on Natural Language Processing for Educational Resources**. Brovets, Bulgaria, pp. 23—29, 2007.

GERMOND, B. From frontier to boundary and back again : the European Union's maritime margins. In: *European Foreign Affairs Review*, v. 15 (1), p. 39-55, 2010.

GIERING, M. E.; ALVES, I. M. R.; MELLO, V. H. D. [orgs.] **Leitura e Produção de Artigo Acadêmico-Científico**. São Leopoldo: UNISINOS, 2010.

GLÄSER, R. The LSP Genre Abstract - Revisited. in **ALSED - Newsletter** , v. 13(4), p. 3-11 1991.

GLÄSER, R. Fachtextsorten im Englischen. **Forum für Fachsprachen-Forschung**, 13. Tübingen, 1990.

GLEDHILL, C., The Discourse Function of Collocation in Research Article Introductions. In: BIBER, D. ; REPPEN, R. [eds.] **Benchmarks in language and linguistics**, v. 1: Lexical Studies, London: Sage Publications, p. 23-45, 2012.

GLEDHILL C. Problems of Adverbial Placement in Learner English and the British National Corpus. In: ALLERTON, D.J. ; TSCHIRHOLD, C. ; WIESER, J. [eds.] **Linguistics, Language Learning and Language Teaching**. (ICSELL 10.) Basel, Schwabe, p. 85-104, 2005.

GLEDHILL, C. The discourse function of collocation in research article introductions. **English for Specific Purposes**, v. 19(2), p. 115-135, 2000.

GLEDHILL, C. The phraseology of rhetoric, collocations and discourse in cancer research abstracts . In: BARRON, C. ; BRUCE, N. [eds.] **Knowledge and Discourse : Proceedings of the International Multidisciplinary Conference, 18-21 June 1996**. Hong Kong : University of Hong Kong, 1999. Disponível em: <<http://ec.hku.hk/kd96proc/authors/papers/Gledhill.htm>>. Acesso em: 26 nov. 2015.

GLEDHILL, C. Collocation and Genre Analysis. The Phraseology of Grammatical Items in Cancer Research Abstracts and Articles. **Zeitschrift für Anglistik und Amerikanistik XLIII**, v.1 (1), p. 11-36, 1995.

GRAETZ, N. Teaching EFL students to extract structural information from abstracts. In: ULIJN, J. M. ; PUGH, A. K. [eds.] **Reading for Professional Purposed: Methods and Materials in Teaching Languages**, Acco, Leuven, Belgium, pages 123–135, 1985.

GRANGER, S. The contibution of learner *corpora* to reference and instructional materials design. In: GRANGER, S. ; GILQUIN, G. ; MEUNIER, F. [eds.] **The Cambridge Handbook of learner Corpus Research**. Cambridge: Cambridge University Press. p. 486-510, 2015.

GRANGER, S. [org.] **Learner English on Computer**. New York: Longman, 1998.

GRANGER, S. The Learner Corpus: A Revolution in Applied Linguistics. **English Today**, v. 39 (10/3), p. 25-29, 1994.

GRANGER, S. ; MEUNIER, F. Phraseology in language learning and teaching: Where to from here? In: MEUNIER, F. ; GRANGER, S. [eds.] **Phraseology in foreign language learning and teaching**. Amsterdam: John Benjamins, p. 247-252, 2008.

GRIES, S. T. Corpus linguistics and theoretical linguistics. A love–hate relationship? Not

necessarily... **International Journal of Corpus Linguistics**, v. 15(3), p. 327–343. John Benjamins Publishing Company, 2010.

GRIES, S. T. What is *Corpus Linguistics*? **Language and Linguistics Compass**, v.3, p. 1–17, 2009.

GRIES, S. T. Phraseology and linguistic theory: a brief survey. In: GRANGER, S. ; HABERT B. ; GRABAR N. ; JACQUEMART P. ; ZWEIGENBAUM P. [eds.] **Building a text corpus for representing the variety of medical language**. Corpus linguistics, *Lancaster*, 2001.

GROOM, N. Pattern and meaning across genres and disciplines: An exploratory study. **English for Academic Purposes**, v. 4, p. 257–277, 2005.

GUIMARÃES, C. A. Structured Abstracts. Narrative Review. **Acta Cirúrgica Brasileira**, v. 21(4), p. 263-268, São Paulo July/Aug. 2006. Disponível em: <<http://dx.doi.org/10.1590/S0102-86502006000400014>> Acesso em: 01 nov. 2015.

GUIMARÃES, E. **A articulação do texto**. 10^a ed. São Paulo: Ática, 2006.

HALLIDAY, M. A. K. Corpus studies and probabilistic grammar. In: AIJMER, K. ; HASAN, B. R. 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.

HALLIDAY, M.A.K. **Spoken and written language**. Oxford: Oxford University Press, 1989.

HALLIDAY, M. A. K.; HASAN, R. **Cohesion in English**. London: Longman, 1978.

HALLIDAY, M.A.K. **Explorations in the functions of language**. London: Edward Arnold, 1973.

HALLIDAY, M. A. K.; MCINTOSH, A.; STREVENSON, P. **The Linguistic Sciences and Language Teaching**. Longmans' Linguistic Library. London: Longmans, 1964.

HARDIE, A. From legacy encodings to Unicode: the graphical and logical principles in the scripts of South Asia. In: **Language Resources and Evaluation**, v. 41 (1), p. 1-25, 2007.

HARDIE, A. Automated part-of-speech analysis of Urdu: conceptual and technical issues. In: YADAVA, Y. ; BHATTARAI, G. ; LOHANI, R. R. ; PRASAIN, B. ; PARAJULI, K. [eds.] **Contemporary issues in Nepalese linguistics**. Kathmandu: Linguistic Society of Nepal, 2005.

HARTMANN, P. Texte als linguistisches Objekt. In: STEMPEL, W. [org.] **Beiträge zur Textlinguistik**. München: Fink, p. 9-29. 1971.

HARWEG, R. **Pronomina und Textkonstitution**. München: Fink, [1968]1979 .

HASAN, R. Text in the Systemic Functional Model. IN: DRESSLER, W. U. **Current Trend in Textlinguistics**. Berlin, New York: Walter di Gruyter, p. 228-246, 1978.

HELLERMAN, J. **Social Actions for Classroom Language Learning**. Clevedon/U.K.: Multilingual Matters, 2008.

HOEUBOECK, A. Some Aspects of Coherence, Genre and Rhetorical Structure and Their

Integration in a Generic Model of Text. **Language Studies Working Papers**, v.1, p. 35-45, 2009.

HOEY, M. **Lexical priming**. London: Routledge, 2005.

HOEY, M. From concordance to text structure: New uses for computer corpora. In: LEWANDOSWKA-TOMASZCZYK, B. ; MELIA, P. J. [org.] **PALC'97 – Practical Applications in Language Corpora**. Lodz: Lodz University Press, p. 229-246, 1997.

HOFFMANN, L. O papel das linguagens especializadas desde meados do século XX. Tradução de ZILIO, L. ; PICKBRENNER, M. In: FINATTO, M. J. B. ; ZILIO, L. [orgs.] **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, p.21-34, 2015. Disponível em: <<http://www.ufrgs.br/ppgletras/>>. Acesso em: 09 Jun. 2015.

HOFFMANN, L. Conceitos Básicos da Linguística de Linguagens Especializadas. Tradução de FINATTO, M. J. B.; KILIAN, C. K. In: FINATTO, M. J. B. ; ZILIO, L. [orgs.] **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, p. 39-50, 2015. Disponível em:<<http://www.ufrgs.br/ppgletras/>>. Acesso em: 10 Jun. 2015.

HOFFMANN, L. Linguagem Especializadas. Tradução de ZILIO, L.; BOCORNY, M. J. F. In: FINATTO, M. J. B. ; ZILIO, L. [orgs.] **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, p.75-80, 2015. Disponível em: <<http://www.ufrgs.br/ppgletras/>>. Acesso em: 10 Jun. 2015.

HOFFMANN, L. Linguagem especializadas como sublinguagens. Tradução de ZILIO, L.; FINATTO, M. J. B. F. In: FINATTO, M. J. B. ; ZILIO, L. [orgs.] **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, p. 89-102, 2015. Disponível em: <<http://www.ufrgs.br/ppgletras/>>. Acesso em: 10 Jun. 2015.

HOFFMANN, L. Gêneros textuais especializados: uma concepção para a formação em línguas estrangeiras voltada para linguagens especializadas. Tradução de KILIAN, C. K.; FINATTO, M. J. B. In: FINATTO, M. J. B.; ZILIO, L. [orgs.] **Textos e Termos por Lothar Hoffmann**. Porto Alegre: Palotti, p. 127-146, 2015. Disponível em: <<http://www.ufrgs.br/ppgletras/>>. Acesso em: 10 Jun. de 2015.

HOFFMANN, L. Conceitos Básico da linguística das linguagens especializadas. Tradução de FINATTO, M. J. B. ; BEVILACQUA, C. R. **Cadernos de Tradução**, p. 79-90, 2004.

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

HOFFMANN, L. **Kommunikationsmittel Fachsprache: eine Einführung**. 2. Aufl. Tübingen: Narr, 1995.

HOFFMANN, L. **Vom Fachwort zum Fachtext: Beitrage zur Angewandten Linguistik**. Tübingen: Gunter Narr Verlag, 1988.

HOFFMANN, L. **Kommunikationsmittel Fachsprache: eine Einführung**. 2. Aufl. Tübingen: Narr. 1985.

HUCKIN, T. N. Abstracting from abstracts. In: HEWINGS, M. [ed.] **Academic writing in context: implications and applications**. Birmingham, UK: University of Birmingham Press, 2001.

HUNSTON, S. **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press, 2002.

HUNSTON, S. A corpus study of some English verbs of attribution. **Functions of Language**, 2, p. 133–158, 1995.

HYLAND, K. *As can be seen: lexical bundles and disciplinary variation*. **English for Specific Purposes**, v. 27, p. 4–21, 2008^a.

HYLAND, K. Academic clusters: text patterning in published and postgraduate writing. **International Journal of Applied Linguistics**, v. 18, n.1, p. 41–61, 2008^b.

HYLAND, K. Disciplinary voices. **Journal of English Text Construction** 1, v.1, Amsterdam: John Benjamins, p. 5-22, 2008^c.

HYLAND, K. Disciplinary differences: language variation in academic discourses. In: HYLAND, K. ; BONDI, M. [eds.] **Academic discourse across disciplines**. Frankfurt: Peter Lan, 2006.

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

HYLAND, K. Genre: language, context, and literacy. **Annual Review of Applied Linguistics**, 22, v. p. 113-135, 2002.

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

ISO 214. (The International Organization for Standardization) **Documentation-Abstracts for Publications and Documentation**, 1976. Disponível em: <<https://www.iso.org/obp/ui/#iso:std:iso:214:ed-1:v1:en>>. Acesso em: 20 out. 2015.

JACOBI, C. C. B. D. **Linguística de Corpus e ensino de espanhol a brasileiros: Descrição de padrões e preparação de atividades didáticas (decir/hablar ; mismo; mientras/en cuanto/ aunque)**. Dissertação de Mestrado, PUCSP, São Paulo, 2001.

JESPERSEN, O. **The philosophy of grammar**. London: George Allen & Unwin, 1924.

JONES, C.; WALLER, D. **Corpus Linguistics for Grammar – A guide for research**. New York: Routledge, 2015.

JORDAN, R. The growth of EAP in Britain. **Journal of English for Academic Purposes**, v. 1(1), p. 69–78, 2002.

KALVERKÄMPER, H. Textuelle Fachsprachen-Linguistik als Aufgabe. **Zeitschrift für Literaturwissenschaft und Linguistik**, v.51/52(13), p.124-166, 1983.

KAN, M. Y.; MCKEOWN, K. R.; KLAVANS, J. L. Domain-specific informative and indicative summarization for information retrieval. **Proceedings of the Document Understanding Workshop (DUC 2001)**. New Orleans. USA: September 2001.

- KENNEDY, G. **An introduction to Corpus Linguistics**. New York: Longman, 1998.
- KENNEDY, C. ; R. BOLITHO. **English for Specific Purposes**. Basingstoke: Macmillan, 1984.
- KILIAN, C. K. **A retomada de unidades de significação especializada em textos de língua alemã e portuguesa sobre gestão de resíduos: uma contribuição para a tradução técnico-científica**. Tese de Doutorado. Universidade Federal do Rio Grande do Sul, 2007.
- KJELLMER, G. A mint of phrases. In: AIJMER, K. ; ALTENBERG, B. [eds.] **English corpus linguistics: Studies in honour of Jan Svartvik**. London: Longman, p. 111-127, 1991.
- KOCH, I. G. V. Trajetória da Linguística Textual. In: **Introdução à lingüística textual**. São Paulo: Martins Fontes, p. 01-46, 2004.
- KOCH, I. G. V. Linguística Textual: Retrospecto e Perspectivas. **Alfa**, São Paulo, v. 41, p. 67-78, 1997.
- KOCH, I. G. V. O texto: construção de sentidos. **Organon**, v. 9 (23), p. 21-27, 1995.
- KRIEGER, M. G. Terminologia: uma entrevista com Maria da Graça Krieger. **ReVEL**, v. 9 (17), 2011. Disponível em: <www.revel.inf.br>. Acesso em: 20 nov. 2015.
- KRIEGER, M. G. Porque Lexicografia e Terminologia: relações textuais? In: FINGER, I. ; COLLISCHONN, G. [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. ISBN: 978-85-7590-115-1, 2008.
- KRIEGER, M. G.; BEVILACQUA, C. R. A pesquisa terminological no Brasil: uma contribuição para a consolidação da área. **Debate Terminológico**, n. 1, 03/2005. Disponível em <http://www.riterm.net/revista/n_1/krieger.pdf>. Acesso em 20 nov. 2015.
- KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à Terminologia – Teoria e Prática**. São Paulo: Contexto, 2004.
- KRIEGER, M. G. ; MACIEL, A. M. B. ; FINATTO, M. J. B. Terminografia das leis do meio ambiente: princípios teórico-metodológicos. **Tradterm**, São Paulo, v. 6, p. 143-169, 2001.
- KRIEGER, M. G. ; SANTIAGO, M. S. ; CABRÉ, M. T. Terminologia em foco: uma entrevista comentada com Maria Teresa Cabré. **Calidoscópio**, v. 11 (3), p. 328-332, set/dez 2013.
- KUPSKE, F. F. A aquisição da linguagem à luz de um paradigma teórica de cognição. **Littera** (UFMA), v. 3, p. 174-190, 2012.
- LABOV, W. Sociolinguistics: an interview with William Labov. **ReVEL**, v.5, n. 9, p. 1-3, 2007.
- LABOV, W. **Sociolinguistic Patterns**. Philadelphia:University of Pennsylvania Press, 1972.
- LABOV, W. **A study of non-standard English**. Washington/DC: ERIC Clearinghouse for Linguistics, 1969.

- LABOV, W. **The Social Stratification of English in New York City**. Cambridge: University Press, 1966.
- LANCASTER, E. W. **Indexing and Abstracting: Theory and Practice**. London: LA Publication, 1991.
- LEA, M. ; STIERER, B. **Student writing in higher education: new contexts**. Buckingham: Society for Research in Higher Education and Open University Press, 2000.
- LEECH, G.; GARSIDE, R.; BRYANT, M. CLAWS4: The tagging of the British National Corpus. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 94)* Kyoto/Japan, p. 622-628, 1994.
- LEECH, G. Corpora and theories of linguistic performance. In: SVARTVIK, J. [org.] *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*, Stockholm, 4-8 August 1991. Berlin, New York: De Gruyter, 1992.
- LEWIS, M. **The Lexical Approach: the state of ELT and a way forward**. Hove: LTP, 1993.
- LUZON MARCO, M. J. Collocational frameworks in medical research papers: a genre-based study. **English for Specific Purposes**, v. 19, p. 63- 86, 2000.
- MANN, W. C.; THOMPSON, S. A. Rhetorical Structure Theory: a theory of text organization. In: POLANYI, L. [ed.] **The Structure of Discourse**, Noewood, N. J.: Ablex, p. 243-281, 1988.
- MARCUSCHI, L. A. Gêneros textuais: configuração, dinamicidade e circulação. In: KARWOSKI, A. M.; GAYDECZKA, B.; BRITO, K. S. [orgs.] **Gêneros textuais: reflexões e ensino**. 2. Ed. Rev e ampl. Rio de Janeiro: Lucerna, p. 23-36, 2006.
- MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. In: DIONÍSIO, A.P.; MACHADO, A. R.; BEZERRA, M. A. [orgs.] **Gêneros textuais & ensino**. Rio de Janeiro: Editora Lucena, p.19-36, 2002.
- MARCUSCHI, L. A. **Aspectos linguísticos, sociais e cognitivos da produção de sentido**, 1998 (mimeo).
- MARCUSCHI, L. A. **Por uma proposta para a classificação dos gêneros textuais**. Recife. Trabalho não publicado, 1996.
- MARCUSCHI, L. A. **Linguística de Texto: o que é e como se faz**. Recife: UFPE, 1983.
- MCENERY, T. ; XIAO, R.; TONO, Y. **Corpus-Based Language Studies**. USA/Canada: Routledge, 2006.
- MCENERY, T. ; XIAO, R.; TONO, Y. Swearing in Modern British English: The Case of Fuck in the BNC. **Language and Literature**, v. 13, p. 235-268, 2004.
- MCENERY, T. ; WILSON, A. **Corpus Linguistics**. Edinburgh: Edinburgh University Press, 1996/2001.
- MCINTOSH, N. Abstract information and structure: experience from an international pediatric meeting. **ESE**, v. 23(1), p. 3-6, 1997.

- MEYER P.G. Statistical text analysis of abstracts: A pilot study on cohesion and schematicity. **Computer Corpora des Englishen**, v. 3, p.17-40, 1988.
- MEUNIER, F. **Phraseology: an interdisciplinary perspective**. Amsterdam: Johns Benjamins, 2008.
- MEUNIER, F.; GRANGER, S. [eds.] **Phraseology in Foreign Language Learning and Teaching**. Amsterdam: John Benjamins Publishing, 2007.
- MEURER, J. L. Esboço de um modelo de produção de textos. In: MEURER, J. L. ; MOTTA-ROTH, D. **Parâmetros de textualização**, Santa Maria: Editora da UFSM, p. 14- 27, 1997.
- MILLER, C. R. Genre as social action. **Quarterly Journal of Speech**, v. 70, p. 151-167, 1984.
- MOON, R. Frequencies and forms of phrasal lexemes in English. In: COWIE, A. P. [ed.] **Phraseology: theory, analysis and applications**. Oxford: Oxford University Press, p. 79-100, 1998.
- MOTTA-ROTH, D. **Rhetorical Features and Disciplinary Cultures: A Genre-Based Study of Academic Book Review in Linguistics, Chemistry and Economics**. Tese de Doutorado. Florianópolis: Programa de Pós-graduação em Inglês/ Universidade Federal de Santa Catarina, 1995.
- MOTTA-ROTH, D. ; HENGES, G. R. **Produção textual na universidade**. São Paulo: Parábola Editorial, 2013.
- MYERS G. **Writing Biology: Texts in the Social Construction of Scientific Knowledge**. Wisconsin: The University of Wisconsin Press, 1990.
- NARITA, M.; KUROKAWA, K.; UTSURO, T. Case Study on the Development of a Computer- Based Support Tool for Assisting Japanese Software Engineers with their English Writing Needs. **IEEE Transactions on Professional Communication**, v. 49 (3), p.194-209, 2003.
- NATTINGER, J. ; DECARRICO, J. **Lexical phrases and language teaching**. New York: Heinle and Heinle, 1992.
- NEIS, I.A. Por uma Gramática Textual. **Letras de Hoje**, v. 44, PUC/RS, p. 21-39, 1981.
- NWOGU K. N. **Discourse variation in medical texts: Schema, theme and cohesion in professional and journalistic accounts**. Unpublished PhD. Thesis, Language Studies Unit, Aston University, 1989.
- O'HALLORAN, K. **How to use corpus linguistics in the study of media discourse**. In: O'KEEFFE, A. ; MCCARTHY, A. ; MCCARTHY, M. [eds.] *The Routledge Handbook of Corpus Linguistics*. London: Routledge, p.563–577, 2010.
- ORASAN, C. Patterns in scientific abstracts. In **Proceedings of Corpus Linguistics, 2001 Conference**. Lancaster: Lancaster University, p. 433-443, 2001.
- PALMER, H. E. **Second interim report on English collocations**. Tokyo: Kaitakusha, 1933.

- PARELMAN, C. **Retóricas**. São Paulo: Martins Fontes, 2004.
- PEARSON, J. **Terms in context**. Amsterdam, Philadelphia: John Benjamins, 1998.
- PINTO, M. ; LANCASTER, F. W. Abstracts and abstracting in knowledge discovery. **Library Trends**, v. 48 (1), p. 234-248, 1999.
- POSSAMAI, V. **Marcadores textuais do artigo científico em comparação português- inglês: um estudo sob a perspectiva da tradução**. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul. Porto Alegre, 2004.
- POSTEGUILLO, S. A genre based approach to the teaching of reading and writing abstracts in Computer Science. In: PIQUÉ, J. ; ANDREU-BESÓ, V.; VIEIRA, D. J. [eds.] **English in Specific Settings**. Valencia: NAU Libres, p. 47-57, 1996.
- POTTS, A.; BEDNAREK, M.; CAPLE, H. How can computer-based methods help researchers to investigate news values in large datasets? A corpus linguistic study of the construction of newsworthiness in the reporting on Hurricane Katrina. **Discourse & Communication**, 9 (2), p. 149-172, February 24, 2015.
- RAJAGOPALAN, K. A Linguística de *Corpus* no tempo e no espaço: visão reflexiva. In: GERBER, R. M. ; VASILÉVSKI, V. **Um percurso para pesquisas com base em corpus**. Florianópolis: Editora da UFSC, 2007.
- RAJAGOPALAN, K. A. O conceito de identidade na linguística: é chegada a hora de uma reconsideração radical? In: SIGNORINI, I. [org.] **Lingua(gem) e identidade**. Campinas: Mercado de Letras, p. 21-45, 1998.
- RAMISCH, C.; VILLAVICENCIO, A. ; KORDONI, V. Introduction to the special issue on multiword expressions. **ACM Transactions on Speech and Language Processing**, v. 10, p. 1-10, 2013.
- RAYSON, P.; ARCHER, D.; PIAO, S. ; MCENERY, A.M. The UCREL semantic analysis system. **Proceedings of the Beyond Named Entity Recognition Semantic Labeling for NLP Tasks Workshop**, Lisbon, Portugal, p. 7-12, 2004.
- RENOUF, A.; SINCLAIR, J. Collocational frameworks in English. In: AIJMER, K.; ALTENBERG, B. [eds.] **English corpus linguistics**. London: Longman, p. 128-143, 1991.
- RICHARDS, J. C.; PLATT, J. **Dictionary of language teaching and applied linguistics**. London: Longman, 1992.
- ROBINSON, P.; ELLIS, N. C. [eds.] **A handbook of cognitive linguistics and second language acquisition**. London: Routledge, 2008.
- ROJO, R. **Multi-letramento na escola**. São Paulo: Parábola, 2012.
- RÖMER, U. The inseparability of lexis and grammar: *Corpus* linguistic perspectives. **Annual Review of Cognitive Linguistics**, v. 7, p. 140-162, 2009.
- RONDEAU, G. **Introduction à la terminologie**. Québec: Gaëtan Morin, 1994.
- ROSCH, E. Principles of Categorization. In: ROSCH, E.; LLOYD, B. [eds.] **Cognition and**

categorization. Hillsdale, NJ: Lawrence Erlbaum, p. 27-48, 1978.

SAGER, J. C. **Curso práctico sobre el procesamiento de la terminología**. Madrid: Pirámide, 1993.

SALAGER-MEYER, F. A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. **English for Specific Purposes**, 11 (2), p. 93-113, 1992.

SALAGER-MEYER, F. Medical English abstracts: How well are they structured? **Journal of the American Society for Information Science, (JASIS)**, v. 42 (7), p. 528-531, 1991.

SALAGER-MEYER, F. Discoursal Flaws In Medical English Abstracts. **Text**, v 10(4), p. 365-384, 1990.

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

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

SALES, R. Teoria Comunicativa da Terminologia (TCT) Como Aporte Teórico para a Representação do Conhecimento Especializado. **VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação**. Salvador: Bahia, 2007.

SAMRAJ, B. An exploration of a genre set: research article abstracts and introductions in two disciplines. **English for Specific Purposes**, v. 24, p.141-156, 2005.

SÁNCHEZ HERNÁNDEZ, P. Lexical bundles in three oral corpora of university students. **Nordic Journal of English Studies**, v.13(1), p.187-209, 2013.

SANTOS, A. R. **Metodologia científica: a construção do conhecimento**. Rio de Janeiro: DP & A Editora, 1999.

SANTOS M. B. D. The textual organization of research paper abstracts in applied linguistics. **Text**, v. 16 (4), p. 481- 499, 1996.

SAUSSURE, F. **Curso de Linguística Geral / Ferdinand de Saussure**. Tradução de: CHELINI, A. PAES, J. P. ; BLIKSTEIN, I. BALLY, C. ; SECHEHAYE, A. [orgs.] ; RIEDLINGER, A. [colab.] São Paulo: Cultrix, 27ª ed. 2006.

SCARAMUCCI, Matilde V. R. Proficiência em LE: Considerações terminológicas e conceituais. **Trabalhos de Linguística Aplicada**, v. 36, p.11-22, 2000.

SCHMITT, N. [ed.] **Formulaic Sequences**. Amsterdam: John Benjamins, 2004.

SCHRÖDER, H. **Subject-oriented texts: languages for special purposes and text theory**. Berlin: De Gruyter, 1991.

SCOTT, M. **Wordsmith Tools 4**. Oxford: Oxford University Press, 1996.

SCOTT, M. ; TRIBBLE, C. **Textual Patterns: key word and corpus analysis in language education**. Amsterdam: John Benjamins, 2006.

SEGRÈ, D. **Cocaina**. Milano: Sonzogno, 1921.

SHEHZAD, W. Announcement of the principal findings and value addition in Computer Science research papers. *Ibérica*, v. 19, p. 97-118, 2010.

SHEPHERD, T. M. G. O Estatuto de Linguística de *Corpus*: Metodologia ou Área da Linguística? *Matraga*, Rio de Janeiro, v.16 (24), p. 150-172, jan./jun., 2009.

SIMPSON-VLACH, R. ; ELLIS, N. An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics*, p. 1-26, 2010.

SINCLAIR, J. The phrase, the whole phrase, and nothing but the phrase. In: GRANGER, S. ; MEUNIER, F. [eds.] **Phraseology: an interdisciplinary perspective**. Amsterdam: Johns, p. 407- 410, 2008.

SINCLAIR, J. **Trust the text: Language, Corpus and Discourse**. London/ New York. Routledge, 2004.

SINCLAIR, J. **Reading Concordances**. UK: Longman, 2003.

SINCLAIR, J. *Corpus Evidence in Language Description*. In: WICHMANN, S. FLIGELSTONE, A. ; MCENERY, T. ; KNOWLES, E. [eds.] **Teaching and Language Corpora**. New York: Addison Wesley Longman, p. 27-39, 1997.

SINCLAIR, J. **Preliminary recommendations on Corpus Typology**. EAGLES Document EAG TCWG CTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale. Unpublished manuscript, 1996. Disponível em: <ftp:// <ftp.ilc.pi.cnr.it>>. Acesso em 04 nov. 2015

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

SINCLAIR, J. **Collins Cobuild English Language Dictionary: Helping Learners with Real English, Revised Edition**. Heinle ELT, 1987.

SINCLAIR, J.M. ; RENOUF, A. A lexical syllabus for language learning. In: CARTER, R. ; MCCARTHY, M. [eds.] **Vocabulary and language teaching**. Harlow: Longman, 1988.

SOARES, M. Letramento e Alfabetização: as muitas facetas. *Revista Brasileira de Educação* v. 25, p. 5-17, 2004.

SOARES, M. **Alfabetização e letramento**. São Paulo: Contexto, 2003.

SOLLACI, L. B. ; PEREIRA, M.G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. *J Med Libr Association*, v. 92 (3), p. 364–37, julho 2004.

STUBBS, N. An example of frequent English phraseology: distributions, structure and functions. In: FACCHINETTI, R. [ed.] **Corpus Linguistics 25 Years on**. New York/Amsterdam: Rodopi, p. 89-106, 2007.

STUBBS, N. Two quantitative methods of studying phraseology in English. **International Journal of Corpus Linguistics**, v. 7(2), p. 215-244, 2002.

STUBBS, M. **Words and Phrases**. Oxford: Blackwell, 2001.

STUBBS, Stubbs, M. British traditions in text analysis: From Firth to Sinclair. In: BAKER,

M. ; FRANCIS, F. ; TOGNINI- BONELLI, E. [eds.] **Text and Technology: In Honour of John Sinclair**. Amsterdam/Philadelphia: John Benjamins, p. 1–46, 1993.

SWALES, J. M. **Aspects of Article Introductions**. Ann Arbor: Michigan University Press, 2011.

SWALES, John. Sobre modelos de análise do discurso. In: BIASI-RODRIGUES, B.; BIASI-RODRIGUES, B.; ARAÚJO, J. C.; SOUSA, C. T. [orgs.] **Gêneros textuais e comunidades discursivas: um diálogo com John Swales**. Belo Horizonte. Editora Autêntica, p.22-46, 2009.

SWALES, J. M. **Research genres: Exploration and applications**. Cambridge: Cambridge University Press, 2004.

SWALES, J. M. Integrated and fragmented worlds: EAP materials and corpus linguistics. In: FLOWERDEW, J. [ed.] **Academic discourse**. Harlow: Longman, p. 150-162, 2002.

SWALES, J. **Other Floors, Other Voices: A Textography of a Small University Building**. New Jersey: Erlbaum, 1998.

SWALES, J. English as Tyrannosaurus rex. **World Englishes**, v. 16, p. 373–82, 1997.

SWALES, J. M. The writing of research articles introduction. **Written Communication**, v. 4 (2), p. 175-191, 1994.

SWALES, J. M. **Genre Analysis. English in Academic & Research Settings**. Cambridge: Cambridge University Press, 1993^a.

SWALES, J. M. **Genre and engagement**. La Revue Belge de la Philologie et L'Histoire. London: Routledge, 1993^b.

SWALES, J. Re-thinking genre: another look at discourse community effects. In: **Re-thinking Genre Colloquium**. Ottawa: Carleton University, 1992 [mimeo].

SWALES, J. M. **Genre Analysis. English in Academic and Research Settings**. Cambridge: Cambridge University Press, 1990.

SWALES, J. M. **Episodes in ESP**. Oxford/New York: Pergamon Institute of English, 1985.

SWALES, J. M. Research into the structures of introductions to journal articles and its applications to the teaching of academic writing. In: WILLIAMS, R. ; SWALES, J. ; KIRKMAN, J. [eds.] **Common Ground: Shared Interests in ESP and Communication Studies**. Oxford: Pergamon Press, ELT Documents 117, p. 77-86, 1984.

SWALES, J. K. **Aspects of article introductions**. Birmingham: Language Studies Unit. University of Aston [Aston ESP Research Reports 1], 1981.

SWALES, J. M. ; FEAKE, C. **Abstracts and the writing of abstracts**. Michigan, University of Michigan Press, 2009.

SWALES, J. M.; FEAKE, C. B. **English in Today's Research World: A Writing Guide**. Ann Arbor: The University of Michigan Press, 2000.

SWALES, J. M.; FEAKE, C. **Academic Writing for Graduate Students: Essential Tasks and Skills (Michigan Series in English for Academic & Professional Purposes)**. Ann

Arbor: The University of Michigan Press, 1994.

TAGNIN, S. E. O. Glossário de Linguística de *Corpus*. In: VIANA, V. ; TAGNIN, S. E. O. **Corpora no Ensino de Línguas Estrangeiras**. São Paulo: HUB Editorial, p. 349-353, 2010.

TEIXEIRA, E. D. Etiquetagem em Linguística de *Corpus*: Possibilidade de Aplicação. In: GERBER, R. M. ; VASILÉSKI, V. [orgs.] **Um percurso para pesquisas com base em corpus**. Florianópolis: Editora da UFSC, p 116-148, 2007.

TEMMERMAN, R. **Towards new ways of terminology description: the sociocognitive-approach**. Amsterdam: John Benjamins, 2000.

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

TEUBERT, W. My version of corpus linguistics. **International Journal of Corpus Linguistics**, v.10 (1), p. 1-13, 2005.

TOGNINI-BONELLI, E. **Corpus linguistics at work**. Amsterdam: John Benjamins, 2001. Pans Mouton, 1981. Disponível em: <ftp:// ftp.ilc.pi.cnr.it.>. Acesso em 05 dez. 2015.

TORRES, L. S. **Escrita científica em português por hispano falantes: recursos linguísticos-computacionais baseados em métodos de alinhamento de textos paralelos**. Tese de doutorado. Universidade de São Paulo. São Carlos, 2015.

VAN DIJK, T. A. **Macrostructures**. Hillsdale: Lawrence Erlbaum, 1980.

VAN DIJK, T. A. Recalling and summarizing complex discourse In: BURGHARDT, W.; HOLKER, K [ed.] **Text Processing Berlin de Gruyter**, p. 49-118, 1979.

VAN DIJK, T.; KINTSCH, W. **Strategies of Discourse Comprehension**. New York: Academic Press, 1983.

VENTOLA, A. Modalization: probability: an explotation into its role in academic writing. In: DUSZAK, A. [ed.] **Culture and Styles in Academic Discourse**. Berlin: Mouton de Gruyter, p. 157-181, 1997.

VIANA, V. ; ZYNGIER, S. ; BARNBROOK, G. [eds.] **Perspectives on Corpus Linguistics**. Amsterdam/ Philadelphia: John Benjamins, p. 47-62, 2011.

VIEIRA, R. **Computational Linguistics – an interview with Renata Vieira**. ReVEL, vol. 2, n. 3, 2004 [www.revel.inf.br]. Translated by Carolina Reolon Jardim. http://www.revel.inf.br/files/entrevistas/revel_3_entrevista_renata_vieira_eng.pdf . Acesso em: 17 de janeiro de 2016.

VILLAVICENCIO, A.; RAMISCH, C. ; MACHADO, A. ; CASELI, H. M. ; FINATTO, Maria Jose B . **Identificação de Expressões Multipalavra em Domínios Específicos**. Linguamática (Braga), v. 2, p. 15-33, 2010.

VILLAVICENCIO, A.; KORDONI, Valia ; ZHANG, Y. ; IDIART, M. ; RAMISCH, C. . **Validation and Evaluation of Automatically Acquired Multiword Expressions for Grammar Engineering**. In: EMNLP-CoNLL 2007, The 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning,

2007, Republica Tcheca. Proceedings of EMNLP-CoNLL 2007, The 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, 2007.

ZANELLA, A. ; HEBERLE, V. *Abstract* ou *summary*: um estudo do metadiscorso e da organização textual de resumos acadêmicos na área biomédica. **Arquivos Catarinenses de Medicina**, v. 34, no. 4, p. 15-22, 2005.

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

ZILIO, L. 2011. Termo E Valor Linguístico: Uma Abordagem Ensaística. **Cadernos do IL**. No. 42, p. 119-128, Junho de 2011. Disponível em: <<http://seer.ufrgs.br/cadernosdoil/article/view/26016>>. Acesso em: 15 nov.2015.

ZILIO, L. Terminologia Textual e *Linguística de Corpus*: estudo em parceria. In: LOPES PERNA, C. B. ; DELGADO KOCH, H. O.; FINATTO, M. J. B. [orgs.] **Linguagens Especializadas em Corpora**. Porto Alegre: EDIPUCRS, p.128-151, 2010.

ZILIO, L : FINATTO, M. J. B. Sobre o autor e sobre o texto Lexicografia de Especialidade e Terminografia. **Cadernos de Tradução** (Porto Alegre), v. 30, p. 47-49, 2012.

WALSH, S. O'KEEFFE, A. Applying *corpus* linguistics and conversation analysis in the investigation of small group teaching in higher education. **Corpus Linguistics and Linguistic Theory**, v. 8(1), p. 159 – 181, 2012.

WEINRICH, H. **Textgrammatik der deutschen Sprache**. Mannheim: Dudenverlag, 1993.

WICHTER, S. **Experten-und Laienwortschätze. Umriss einer Lexikologie der Vertikalität**. Tübingen: Niemeyer, 1994.

WILLIAMS, G. La linguistique de corpus: Une affaire prépositionnelle. In: RASTIER, F. ; BALLABRIGA, M. Ballabriga [eds.] **Corpus en Lettres et Sciences Sociales: Des Documents Numériques à l'Interprétation**. Paris: Texto, p. 151–158, 2006.

WOOD, A. International scientific English. In: FLOWERDEW, J.; PEACOCK, M. [eds.] **Research perspectives on English for Academic Purposes**. Cambridge, UK: Cambridge University Press, p. 71-83, 2001.

WRAY, A. **Formulaic language: Pushing the boundaries**. Oxford: Oxford University Press, 2008.

WRAY, A. Formulaic sequences in second language teaching: Principle and Practice. **Applied Linguistics**, v. 2 (4), p. 487-489, 2000.

XIAO, R. ; MCENERY, T. Collocation. Semantic Prosody and Near Synonymy: A Cross-Linguistic Perspective. **Applied Linguistics**, v. 27 (1), p. 103–129, 2006.

10. ANEXOS

I - Pacotes lexicais do *subcorpus* internacional em ordem de frequência e distribuição

Tabela (20): Corpus CISA – SubCorpus Internacional – Pacotes Lexicais

Sequência	Frequência	Distribuição	Ocorrência
1	23	16	<i>with the use of</i>
2	19	15	<i>was associated with a</i>
3	14	10	<i>associated with a lower</i>
4	13	11	<i>had no affect on</i>
5	13	13	<i>the objective was to</i>
6	12	12	<i>body mass index bmi</i>
7	12	12	<i>of base case analysis</i>
8	12	10	<i>on the basis of</i>
9	12	12	<i>results of base case</i>
10	12	8	<i>were associated with a</i>
11	12	11	<i>were randomly assigned to</i>
12	11	10	<i>for the treatment of</i>
13	11	11	<i>results of sensitivity analysis</i>
14	11	10	<i>the American college of</i>
15	10	6	<i>the cost effectiveness of</i>
16	9	9	<i>of this study was</i>
17	9	9	<i>this study was to</i>
18	9	7	<i>was associated with higher</i>
19	9	6	<i>with a lower risk</i>
20	8	6	<i>a lower risk of</i>
21	8	8	<i>American college of physician</i>
22	8	7	<i>benefits and harms of</i>
23	8	8	<i>central register of controlled</i>

24	8	8	<i>cochrane central register of</i>
25	8	7	<i>in this study we</i>
26	8	8	<i>preventive services task force</i>
27	8	8	<i>register of controlled trials</i>
28	8	8	<i>the use of a</i>
29	8	6	<i>was associated with lower</i>
30	7	7	<i>and body mass index</i>
31	7	7	<i>assessed risk of bias</i>
32	7	7	<i>assessment of insulin resistance</i>
33	7	5	<i>associated with increased risk</i>
34	7	5	<i>in the presence of</i>
35	7	7	<i>incremental cost effectiveness ratios</i>
36	7	5	<i>intake was associated with</i>
37	7	7	<i>model assessment of insulin</i>
38	7	6	<i>quality adjusted life years</i>
39	7	7	<i>randomized controlled trials rcts</i>
40	7	7	<i>studies are needed to</i>
41	7	7	<i>the aim was to</i>
42	7	7	<i>these results suggest that</i>
43	7	7	<i>was not associated with</i>
44	7	7	<i>years of follow up</i>
45	6	6	<i>a wide range of</i>
46	6	6	<i>an important role in</i>
47	6	6	<i>body mass index in</i>
48	6	5	<i>compared with the lowest</i>
49	6	6	<i>cox proportional hazards models</i>
50	6	6	<i>hepatitis c virus hcv</i>
51	6	6	<i>homeostasis model assessment of</i>
52	6	6	<i>in vitro and in</i>

53	6	6	<i>investigated the effects of</i>
54	6	6	<i>is associated with a</i>
55	6	6	<i>little is known about</i>
56	6	6	<i>mass index in kg</i>
57	6	6	<i>reactive oxygen species ros</i>
58	6	6	<i>regression was used to</i>
59	6	6	<i>selection randomized controlled trials</i>
60	6	6	<i>study selection randomized controlled</i>
61	6	6	<i>the primary outcome was</i>
62	6	6	<i>trial was registered at</i>
63	6	6	<i>was inversely associated with</i>
64	6	6	<i>was to examine the</i>
65	5	5	<i>aim of this study</i>
66	5	5	<i>and reference lists study</i>
67	5	5	<i>and strength of evidence</i>
68	5	5	<i>base case analysis the</i>
69	5	5	<i>c virus hcv infection</i>
70	5	5	<i>estimate the cost effectiveness</i>
71	5	5	<i>european prospective investigation into</i>
72	5	5	<i>examine the association between</i>
73	5	5	<i>has been associated with</i>
74	5	5	<i>have been associated with</i>
75	5	5	<i>hazards models were used</i>
76	5	5	<i>in the absence of</i>
77	5	5	<i>in this review we</i>
78	5	5	<i>models were used to</i>
79	5	5	<i>of bias data synthesis</i>
80	5	5	<i>on the association between</i>
81	5	5	<i>person years of follow</i>

82	5	5	<i>plays an important role</i>
83	5	5	<i>proportional hazards models were</i>
84	5	5	<i>prospective investigation into cancer</i>
85	5	5	<i>tested the hypothesis that</i>
86	5	5	<i>the aim of this</i>
87	5	5	<i>the present study was</i>
88	5	5	<i>these data suggest that</i>
89	5	5	<i>this trial was registered</i>
90	5	5	<i>to estimate the cost</i>
91	5	5	<i>to examine the association</i>
92	5	5	<i>used to evaluate the</i>
93	5	5	<i>vitro and in vivo</i>
94	5	5	<i>was assessed by using</i>
95	5	5	<i>was to investigate the</i>
96	5	5	<i>were estimated by using</i>

Fonte: Freitas (2016)

II - Pacotes lexicais do *subcorpus* brasileiro em ordem de frequência e distribuição

Tabela (21): Corpus CISA – SubCorpus Brasileiro – Pacotes Lexicais

Sequência	Frequência	Distribuição	Ocorrência
1	37	36	<i>of this study was</i>
2	36	36	<i>this study was to</i>
3	30	28	<i>the aim of this</i>
4	24	23	<i>aim of this study</i>
5	22	21	<i>as well as the</i>
6	19	16	<i>in relation to the</i>
7	19	19	<i>was to evaluate the</i>
8	17	17	<i>study was to evaluate</i>

9	16	16	<i>the objective of this</i>
10	13	12	<i>on the other hand</i>
11	13	13	<i>this study aimed to</i>
12	12	7	<i>at the end of</i>
13	12	10	<i>it was observed that</i>
14	12	12	<i>objective of this study</i>
15	12	12	<i>the analysis of the</i>
16	11	11	<i>of the present study</i>
17	11	8	<i>the nutritional status of</i>
18	11	11	<i>the present study was</i>
19	10	9	<i>for the treatment of</i>
20	10	10	<i>is one of the</i>
21	10	6	<i>the end of the</i>
22	9	8	<i>in the presence of</i>
23	9	9	<i>of this study is</i>
24	9	9	<i>present study was to</i>
25	9	6	<i>the quality of life</i>
26	9	9	<i>these results suggest that</i>
27	8	7	<i>an increase in the</i>
28	8	8	<i>body mass index bmi</i>
29	8	8	<i>evaluate the effect of</i>
30	8	7	<i>it was found that</i>
31	8	8	<i>results of this study</i>
32	8	8	<i>the results of this</i>
33	8	8	<i>to evaluate the effect</i>
34	8	8	<i>was observed that the</i>
35	8	8	<i>was to analyze the</i>
36	7	7	<i>in the field of</i>
37	7	7	<i>in the state of</i>

38	7	7	<i>in this study the</i>
39	7	6	<i>in this study we</i>
40	7	6	<i>it is necessary to</i>
41	7	7	<i>it was used a* (sic)</i>
42	7	7	<i>one of the most</i>
43	7	5	<i>quality of life of</i>
44	7	7	<i>study was to investigate</i>
45	7	6	<i>the family health strategy</i>
46	7	7	<i>the purpose of this</i>
47	7	6	<i>the use of the</i>
48	7	6	<i>was observed in the</i>
49	6	5	<i>a cross sectional study</i>
50	6	6	<i>aimed to evaluate the</i>
51	6	5	<i>for the development of</i>
52	6	6	<i>in the city of</i>
53	6	6	<i>in the process of</i>
54	6	6	<i>it is possible to</i>
55	6	6	<i>results suggest that the</i>
56	6	6	<i>study was to analyze</i>
57	6	6	<i>the beginning of the</i>
58	6	5	<i>the objective was to</i>
59	6	6	<i>the other hand the</i>
60	6	6	<i>the results of the</i>
61	6	6	<i>the university hospital of</i>
62	6	6	<i>therefore this study aimed</i>
63	6	6	<i>this study is to</i>
64	6	6	<i>was to investigate the</i>
65	6	6	<i>when compared to the</i>
66	5	5	<i>as a result of</i>

67	5	5	<i>can be used as</i>
68	5	5	<i>cross sectional study with</i>
69	5	5	<i>in children and adolescents</i>
70	5	5	<i>in the present work* (sic)</i>
71	5	5	<i>it is known that</i>
72	5	5	<i>it was concluded that</i>
73	5	5	<i>mass index bmi and</i>
74	5	5	<i>objective of the present</i>
75	5	5	<i>of this work was</i>
76	5	5	<i>results indicate that the</i>
77	5	5	<i>study aimed to evaluate</i>
78	5	5	<i>study was carried out</i>
79	5	5	<i>the aim of the</i>
80	5	5	<i>the effect of the</i>
81	5	5	<i>the goal of this</i>
82	5	5	<i>the world health organization</i>
83	5	5	<i>this work was to* (sic)</i>
84	5	5	<i>to the presence of</i>
85	5	5	<i>was able to reduce</i>
86	5	5	<i>was evaluated using the</i>
87	5	5	<i>were not associated with</i>
88	5	5	<i>were used in the</i>

Fonte: Freitas (2016)