

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP

Ana Julia Perrotti-Garcia

**Artigos Médicos em Inglês, Publicados em Periódicos do Brasil e
do Exterior: uma análise a partir de *corpora* comparáveis**

Mestrado em Linguística Aplicada e Estudos da Linguagem

São Paulo
2009

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO
PUC-SP

Ana Julia Perrotti-Garcia

**Artigos Médicos em Inglês, Publicados em Periódicos do Brasil e
do Exterior: uma análise a partir de *corpora* comparáveis**

Mestrado em Linguística Aplicada e Estudos da Linguagem

**Dissertação apresentada à Banca
Examinadora da Pontifícia Universidade
Católica de São Paulo, como exigência parcial
para a obtenção do título de mestre em
Linguística Aplicada e Estudos da
Linguagem, sob orientação do Prof. Dr.
Antonio Paulo Berber Sardinha**

São Paulo

2009

FOLHA DE APROVAÇÃO

BANCA EXAMINADORA

AUTORIZAÇÃO

Na qualidade de autora, autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução parcial ou total desta dissertação por processos fotocopiadores ou eletrônicos

São Paulo, maio de 2009

Para meu pai, em sua ausência cada vez mais presente.

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus, pela luz infinita e pelos momentos em que tem me carregado no colo, e aos amigos invisíveis que se fazem presentes nos momentos mais necessários.

Agradeço ao Sergio, companheiro de todas as horas, pelo apoio incondicional e pela ajuda constante e a meus filhos Julianna e Francesco, por compreenderem meus silêncios e distanciamentos.

Agradeço ao Prof. Dr. Tony Berber Sardinha, meu orientador, mestre e fonte de inspiração, cuja inteligência, perseverança, generosidade e competência unem-se para compor um ser humano especial.

À querida amiga e colega Dra. Maria Inês Otranto e ao Dr. Sergio Altino Franzi, que participaram da banca de qualificação, e tanto contribuíram para o aprimoramento deste trabalho.

À Prof^a. Dr^a. Tania Maria Granja Shepherd, que encantou com os *clusters*.

Aos professores do LAEL, por compartilharem seus conhecimentos. Em especial às professoras que, durante esta jornada, tornaram-se grandes amigas e luzes para novos caminhos: Cecilinha, Sumiko e Mara.

Aos colegas orientandos do Tony, por suas sugestões, amizade e companheirismo.

Aos demais colegas do LAEL, que conheci em minhas andanças por outras disciplinas ofertadas pelo Programa, pela simpatia e coleguismo.

Aos funcionários da PUC e do LAEL, por sua presença valiosa, necessária e transparente.

Ao aluno Julio Rolim Machado Santos, pela ajuda com os gráficos.

À CAPES, que contribuiu com o suporte financeiro providencial.

Ao Sergio, por sua prontidão e revisão minuciosa.

A todos aqueles que, direta ou indiretamente, tenham contribuído para a concretização deste trabalho.

Caminante, son tus huellas
el camino y nada más;
caminante, no hay camino,
se hace camino al andar.

Al andar se hace camino
y al volver la vista atrás
se ve la senda que nunca
se ha de volver a pisar.

Caminante no hay camino
sino estelas en la mar...

Antonio Machado

“I was like a boy playing on the sea-shore,
and diverting myself now and then finding a smoother pebble
or a prettier shell than ordinary,
whilst the great ocean of truth lay all undiscovered before me.”

Isaac Newton

RESUMO

As revistas médicas brasileiras têm apresentado, nos últimos anos, um número crescente de artigos publicados em inglês, em um esforço para atrair leitores do exterior e dar maior visibilidade aos textos de pesquisadores nacionais. Entretanto, pouco se sabe sobre as características da língua inglesa empregada nesses textos. Dois corpora comparáveis, com aproximadamente cinco milhões de *tokens* cada foram especialmente coletados para o presente estudo. Ambos contêm artigos de revistas médicas escritos em inglês. O *corpus* BRAZIL com artigos publicados em revistas médicas brasileiras e o *corpus* ABROAD com artigos publicados em revistas médicas dos Estados Unidos e Europa, tanto em inglês norte-americano quanto britânico. Após analisar a língua usada em ambos os *corpora*, usando ferramentas do programa *WordSmith* (*WordList*, *clusters* e o concordanceador *Concord*) e uma rotina de computador desenvolvida em linguagem Shell, detectamos que o verbo *submit* teve 1.111 ocorrências no *corpus* BRAZIL e apenas quatro no *corpus* ABROAD. O objetivo específico deste estudo foi determinar como os médicos brasileiros usaram o verbo *submit*, uma tradução *prima facie* da palavra “*submit*”, e detectar quais estruturas gramaticais foram usadas no *corpus* ABROAD para expressar o mesmo significado. As estruturas encontradas no *corpus* ABROAD para expressar a ideia de “*submit*-se a um exame, teste ou terapia” foram: uso de verbos (tais como *undergo*; *receive*; *take*; *have*; *initiate*, *start*, *begin*, *use*, *continue*, *allocate*, *assign*, *attend*, *complete*; *to be taken*, *be referred for*; *perform*, *do*, *be considered for*, *eligible for*, *to be given*, *to be subjected to*, *subject to*) em diferentes tempos, vozes e modos; uso de adjetivação; uso de preposições (*after*, *on*, *with*, *without*).

Palavras-chave: Linguística de *Corpus*; tradução; *clusters*; revistas médicas

ABSTRACT

In recent years, Brazilian medical journals have increased the number of articles published in English in their issues, in an effort to attract readers from outside Brazil. Little is known about the quality of English language used in these papers, however. Two comparable corpora of approximately 5 million tokens each were collected especially for this study and contain full papers from medical journals. Corpus BRAZIL contains full papers from medical journals published in Brazil and corpus ABROAD contains full papers published in American and European medical journals, both in British English and American English. After analyzing the language used in both corpora, using WordSmith Tools (WordList, clusters; and Concord, and a script in Unix shell language, we detected that the verb submit had 1,111 occurrences in corpus BRAZIL and only 4 in corpus ABROAD. We then looked more closely at the verb “submit”, a *prima facie* translation of the Portuguese word “*submeter*”, in order to detect which grammatical structures were used in corpus ABROAD to express the same meaning. The strategies found in corpus ABROAD to express the idea studied were: use of verbs (such as *undergo; receive; take; have; initiate, start, begin, use, continue, allocate, assign, attend, complete; to be taken, be referred for; perform, do, be considered for, eligible for, to be given, to be subjected to, subject to*) in different tenses, voices and modes; use of adjectives; use of prepositions (*after, on, with, without*).

Key-words: Corpus Linguistics; translation; clusters; medical journals

Lista de Ilustrações

	QUADROS	página
1	Ferramentas, utilitários, instrumentos e funções disponíveis no <i>WordSmith Tools</i>	48
2	Padrões apresentados no <i>corpus</i> BRAZIL pela análise de todas as ocorrências do trigrama <i>were submitted to</i> nas primeiras 25 linhas de concordância	57
3	Padrões apresentados no <i>corpus</i> BRAZIL pela análise de todas as ocorrências do trigrama <i>was submitted to</i> nas primeiras 25 linhas de concordância	58
4	Padrões apresentados no <i>corpus</i> BRAZIL pela análise de todas as ocorrências do trigrama <i>patients submitted to</i> nas primeiras 25 linhas de concordância	59
5	<i>Corpus</i> ABROAD: parâmetros de pesquisa das linhas de concordância	62
	GRÁFICOS	
1	Escolhas usadas no <i>corpus</i> ABROAD para expressar a ideia de “submeter-se”.	101

Lista de Tabelas

	TABELAS	página
1	Revistas científicas que compuseram o <i>corpus</i> de estudo (BRAZIL, com revistas brasileiras)	42
2	Revistas científicas que compuseram o <i>subcorpus</i> de referência com revistas médicas publicadas no exterior, com artigos em inglês norte-americano [<i>American English</i>].	42
3	Revistas científicas que compuseram o <i>subcorpus</i> de referência com revistas médicas publicadas no exterior, com artigos em inglês britânico [<i>British English</i>].	43
4	Temas abordados pelas revistas científicas das quais foram coletados os textos para compor os corpora.	44
5	Trigramas exclusivos do <i>corpus</i> BRAZIL, primeiros 25 resultados obtidos.	55
6	Resumo esquemático dos resultados da busca de <i>patient*</i> e <i>*therapy</i> simultaneamente, no <i>corpus</i> ABROAD.	77
7	Resumo esquemático dos resultados da busca de <i>patient*</i> e <i>*ectomy</i> simultaneamente, no <i>corpus</i> ABROAD.	85
8	Resumo esquemático dos resultados da busca de <i>patient*</i> e <i>procedure*</i> simultaneamente, no <i>corpus</i> ABROAD.	89
9	Resumo esquemático dos resultados da busca de <i>patient*</i> e <i>*scopy</i> simultaneamente, no <i>corpus</i> ABROAD.	94
10	Resumo esquemático das escolhas empregadas pelos autores de textos do <i>corpus</i> ABROAD para expressar o sentido de <i>submit</i> através do uso de sintagmas nominais.	98
11	Resumo esquemático das escolhas do <i>corpus</i> ABROAD para expressar o sentido de <i>submit</i> através do uso de diferentes sintagmas verbais.	99

Lista de Abreviaturas e Siglas

WST	<i>WordSmith Tools</i>
LA	Linguística Aplicada
LC	Linguística de <i>Corpus</i>
EUA	Estados Unidos
ESP	<i>English for Specific Purposes</i>
EAP	<i>English for Academic Purposes</i>

SUMÁRIO

INTRODUÇÃO	13
1 Relevância da pesquisa e justificativa	13
2 Objetivos e questões de pesquisa.....	18
3 Delimitação do campo de pesquisa.....	19
4 Organização da dissertação.....	22
CAPÍTULO 1. Fundamentação Teórica.....	24
1.1 A Linguística de <i>Corpus</i>	24
1.1.1 Terminologia relacionada à Linguística de <i>Corpus</i>	26
CAPÍTULO 2 Metodologia.....	37
2.1 Objetivos da pesquisa.....	37
2.2 Escolha da área de pesquisa.....	38
2.3 Seleção dos <i>corpora</i>	39
2.3.1 Critérios para a montagem dos <i>corpora</i>	39
2.3.2 Perfil dos <i>corpora</i> de estudo e de referência.....	41
2.3.3 Problemas enfrentados na definição dos textos que compuseram os <i>corpora</i>	44
2.4 Ferramentas computacionais usadas na preparação e análise dos dados.....	46
2.4.1 O programa <i>WordSmith Tools</i>	46
2.5 Definição da extensão dos <i>clusters</i>	49
2.6 Procedimentos metodológicos gerais da pesquisa	50
2.7 Convenções adotadas para a apresentação dos padrões	51
CAPÍTULO 3. Análise de Dados.....	52
3.1 Trigramas mais frequentes no <i>corpus</i> de referência.....	53
3.2 Trigramas mais frequentes no <i>corpus</i> de estudo.....	53
3.3 Trigramas exclusivos do <i>corpus</i> BRAZIL.....	54
3.4 Análise dos trigramas contendo o verbo <i>submit</i>	56
3.5 Retextualização: padrões equivalentes a <i>submitted</i>	60
3.6 Discussão sobre os dados obtidos.....	95
CONSIDERAÇÕES FINAIS.....	105
REFERÊNCIAS.....	108
APÊNDICES.....	118

INTRODUÇÃO

1. Relevância da pesquisa e justificativa

A produção científica brasileira na área médica apresentou uma crescente internacionalização, além de um aumento expressivo nos últimos 25 anos. Há três aspectos importantes a respeito da questão da publicação de artigos médicos em revistas científicas. O primeiro refere-se ao aumento da produção científica dos autores brasileiros. Ao avaliar os dados sobre os trabalhos com a participação de autores brasileiros, indexados na base Pascal (uma base bibliográfica internacional e multidisciplinar francesa que cobre parte significativa da produção científica nacional [FARIA, QUONIAM & MUGNAINI, 2002]), conforme o ano de publicação observa-se que, entre 1983 e 1990 foram publicados 5.004. Nos dois anos seguintes (1992 e 1993) houve um total de 5.837 publicações de autores brasileiros na referida base. A partir de 1996, o número de publicações foi crescendo ano a ano, sendo de 5.822 artigos publicados em 2000. (MUGNAINI, JANNUZZI & QUONIM, 2004)

O segundo aspecto refere-se à tendência de os autores brasileiros publicarem em revistas internacionais. Em relação à essa tendência, entre 1983 e 1990, mais da metade dos trabalhos de pesquisadores de instituições brasileiras indexados na base Pascal foram publicados em revistas do Brasil; ao final da década de 90, menos de 10% dos artigos de brasileiros foram publicados em revistas nacionais. (MUGNAINI, JANNUZZI & QUONIM, 2004)

O terceiro aspecto importante relativo à questão da publicação de artigos médicos em periódicos internacionais refere-se às características da linguagem empregada ao redigir esses artigos. Ao elaborar um artigo científico, o pesquisador procura comunicar à comunidade científica os resultados de sua pesquisa. Para tanto, precisa lançar mão de uma linguagem “adequada à comunicação com seus pares” (POBLACIÓN *et al.*, 2002, p. 359). Población *et al.* (2003, p.501) afirmam que as publicações nacionais brevemente alcançarão um padrão de qualidade que lhes trará

visibilidade e um fator de impacto compatível com os periódicos internacionais da área. Com relação à internacionalização, para que as revistas científicas brasileiras tenham abrangência internacional, além de fatores ligados ao periódico em si (como periodicidade adequada; regularidade de edição; publicação impressa [*print*] e eletrônica [*online*], entre outros), é necessário que os artigos sejam elaborados segundo as diretrizes aceitas internacionalmente (como, por exemplo, as diretrizes do *Council of Biology Editors; International Committee of Medical Journal Editors [1987]*, UNESCO/INISIST [1979] e da *American Medical Association*, entre outras) e que sejam redigidos em um “bom inglês” (DÖBEREINER, 2006).

A noção de “bom inglês”, segundo análise nossa, refere-se à língua inglesa escrita com padrões léxico-gramaticais que estejam em proximidade idiomática com um padrão de prestígio aceito pela comunidade científica. Esse padrão, para a comunidade médica, pode ser entendido como aquele posto em prática nos artigos publicados em revistas publicadas nos Estados Unidos e Europa (segundo observações empíricas feitas ao longo de nossa experiência profissional).

Em um esforço de ampliar o acesso ao seu conteúdo, muitas revistas médicas publicadas no Brasil vêm sendo redigidas em português e em inglês. Entretanto, pouco se sabe sobre as características do inglês apresentado por esses periódicos. Será que estas revistas científicas brasileiras publicadas em inglês realmente apresentam padrões léxico-gramaticais que as colocam em igualdade de condições com os textos publicados nas revistas científicas de referência internacional? A partir dessa inquietação, procuramos comparar, do ponto de vista da Linguística de *Corpus*, a linguagem dos textos médicos publicados em língua inglesa no Brasil, com textos médicos publicados em inglês em periódicos similares, lançados na América do Norte e na Europa.

Em diversos países do mundo, muitos pesquisadores já se ocuparam da análise da produção científica a partir da avaliação qualitativa do conteúdo científico dos textos médicos (SARMENTO E SOUZA, GREGÓRIO & FORETI, 2004) e da análise quantitativa dos textos publicados em periódicos de diferentes áreas (CASTRO, FERREIRA & VIDILI, 1996;

MIRANDA & PEREIRA, 1996; SPINAK 1996; LOPEZ-CÓZAR 1997; PEREZ GÓMEZ, 1998; WORMELL, 1998). As instruções dadas aos autores também já foram estudadas (GOLDENBERG *et col.*, 1995; SOSA-DE-MARTINEZ, CARNEVALE & AVENDAÑO-INESTRILLAS, 1995; SCHRIGER *et al.*, 2003; SORINOLA *et al.*, 2004), uma vez que o autor que vai publicar em uma revista científica deve seguir as normas de formatação de redação do periódico onde pretende a publicação.

Segundo Pellizzon *et al.* (2007, p. 504), o prestígio de uma revista científica está “atrelado ao rigor ético e científico com que seus artigos são produzidos”. Além disso, a qualidade dos seus textos também é importante, pois não existe rigor ético nem científico sem um texto claro e compreensível onde isso possa ser verificado. Também existem outros fatores que influenciam no aumento do número de leitores de um artigo científico. O idioma da publicação é um elemento que está relacionado com a divulgabilidade dos textos publicados, uma vez que o inglês é considerado “idioma científico internacional”. (MUGNAINI, JANNUZZI & QUONIAM, 2004)

Ao estudarmos as publicações científicas nacionais, observa-se que de 20% a 25% do total são artigos publicados em revistas indexadas nas bases de dados internacionais (MENEHINI, 1998). A indexação dos periódicos em bases de dados propicia que a informação se torne visível à comunidade científica de forma rápida e sistemática (BRAILE, BRANDAU & MONTEIRO, 2007) e “significa reconhecimento de mérito, aval à qualidade de seus artigos” (RIBEIRO, 2006, Editorial). Por outro lado, os artigos publicados em periódicos não indexados correspondem a cerca de 80% do total. Assim, a imagem de um *iceberg* parece ser adequada para representar nossa produção científica: com uma parte visível correspondendo à minoria dos textos publicados e o restante, pouco visível, que se mantém submerso (MENEHINI, 1998). É importante enfatizar que, para o presente estudo, avaliamos apenas as revistas brasileiras indexadas nas bases de dados internacionais, devido à maior visibilidade de seu conteúdo para a comunidade internacional e por supormos que seus textos, se redigidos em inglês, deverão estar de acordo com as exigências de publicação e qualidade dos demais periódicos indexados nessas bases – de onde também coletamos os textos para compor o *corpus* de referência.

Na área de Ciências da Saúde, que foi o objeto de nossa pesquisa, publicar é uma necessidade, tanto para os pesquisadores quanto para a comunidade científica. Muitos pesquisadores brasileiros têm necessidade de redigir um artigo médico em inglês, para poder divulgar seus trabalhos científicos. Para o autor, as vantagens de ter um artigo publicado estão relacionadas ao reconhecimento de seu “esforço intelectual, estabelecimento e sedimentação de sua reputação de pesquisador por meio de acreditação pública, garantia de continuidade de seus projetos, prestígio e obtenção de posições acadêmicas hierarquicamente superiores” (BENNETT & TAYLOR, 2003). Parte dessas vantagens, obviamente, relaciona-se a fatores econômicos, uma vez que a publicação de artigos científicos é um fator levado em consideração para os pesquisadores que buscam patrocínio para seus projetos. A ponto de ter sido cunhado o termo “*publish or perish*”, ou seja, “publique ou perecerá” (tradução nossa), que relaciona a maior produtividade científica diretamente ao sucesso acadêmico entre os pesquisadores da área da saúde. Entretanto, tal opinião tem sido questionada por alguns autores (RELMAN, 1977; ANGELL, 1986; COELHO *et col.*, 2003) que consideram bastante complexo o julgamento da produção científica, o qual vai além do número de publicações (MONTEIRO *et col.*, 2004).

Ao determinarmos os padrões léxico-gramaticais presentes nos textos publicados em inglês nas revistas médicas do exterior, comparando-os com os padrões encontrados nas revistas médicas publicadas em inglês no Brasil, podemos apontar quais aspectos precisariam ser modificados na redação dos artigos em inglês das revistas brasileiras, e as habilidades que os autores deveriam desenvolver para que pudessem fazer escolhas mais próximas daquelas encontradas nos artigos que correspondem a um padrão de prestígio, assemelhando-se aos publicados nas revistas médicas dos países anglófonos. Assim, secundariamente, mas de igual relevância, o presente trabalho busca apontar idiosincrasias recorrentes nos textos publicados em revistas nacionais, escritos em inglês. A partir desses itens detectados, em uma etapa futura, outros estudos poderão estruturar abordagens didáticas que venham a preencher essas lacunas de aprendizado, caso existam.

Por fim, a relevância da pesquisa também se estende a outras áreas do saber. Como os textos coletados são basicamente da área de Ciências da Saúde, mas analisados com uma abordagem linguística, tanto profissionais de Medicina, Odontologia, Farmacologia, enfim, da área da saúde, em geral, quanto tradutores e intérpretes talvez possam se beneficiar dos resultados do presente estudo, caso sejam levados a eles de modo relevante para seus contextos, no futuro.

A escolha de artigos de Medicina foi motivada inicialmente pelo fato de já termos um conhecimento aprofundado sobre o assunto, por nossa graduação em Odontologia e especialização em Cirurgia e Traumatologia Bucomaxilofacial, e por termos um envolvimento razoável com a terminologia e a produção de textos nessa área, dado nosso trabalho de mais de 15 anos com tradução e ensino de inglês médico e odontológico.

Os periódicos dos quais foram coletados os artigos científicos para compor o *corpus* de artigos científicos publicados em periódicos brasileiros (por nós denominado *corpus* BRAZIL) são todos de acesso livre pelo banco de dados *SciELO (Scientific Electronic Library Online)* e *DEDALUS* (Banco de Dados Bibliográficos da USP) e são publicados por associações científicas, instituições de ensino superior e órgãos de classe de qualidade e valor reconhecidos no meio médico brasileiro e internacional.

O sistema de classificação de periódicos Qualis foi criado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), do Ministério da Educação, para a estratificação da qualidade da produção intelectual dos programas de pós-graduação. Esse sistema atribui uma letra (de A a C, em ordem decrescente de importância) aos periódicos avaliados. Os periódicos dos quais foram coletados os artigos científicos para compor o *corpus* BRAZIL foram todos classificados como Qualis A ou B (a relação completa dos periódicos e das entidadesadoras, assim como sua classificação QUALIS, encontram-se no Apêndice A).

Os periódicos dos quais foram coletados os artigos científicos para compor o *corpus* de textos publicados em revistas científicas do exterior (por nós denominado *corpus* ABROAD) foram disponibilizados pela *National Library of Medicine* dos EUA e são todos de acesso pela *PubMed* e *DOAJ*

(*Directory of Open Access Journals*). Além disso, todos os periódicos (exceto dois) têm seus artigos selecionados pelo sistema de revisão por pares (*peer review*), considerado no meio médico como a maneira mais adequada de garantir a qualidade científica das publicações. Apenas os periódicos *Clinical Diabetes* e *American Journal of Otolaryngology: Head Neck Medicine and Surgery* têm seus artigos selecionados por um conselho editorial. A relação completa dos periódicos e das entidades publicadoras encontra-se no Apêndice B.

2 Objetivos e questões de pesquisa

A pesquisa descrita aqui teve como objetivo geral identificar os padrões léxico-gramaticais característicos de textos médicos escritos em inglês publicados em revistas do Brasil e, a seguir, procurar determinar se tais padrões são recorrentes em textos de mesmo tipo textual (artigos científicos), registro (aqui entendido como o nível de formalidade do texto) e público-alvo (profissionais de saúde), publicados nos Estados Unidos e Europa. Para tanto, foram utilizadas as teorias e as ferramentas da Linguística de *Corpus*.

Acreditamos que os trabalhos científicos publicados em inglês em revistas editadas em países anglófonos que compuseram nos *corpus* de referência possam ser considerados como uma base de referência para comparação, pois seguem diretrizes rígidas de redação e também por havermos escolhido apenas revistas científicas indexadas nas bases de dados PubMed (um serviço da *U.S. National Library of Medicine*) e DOAJ (*Directory of Open Access Journals*) que apenas aceitam incluir em sua base de dados revistas científicas cujos artigos sejam revistos por pares (*peer review*) e com controle de qualidade editorial. Os artigos de revistas nacionais foram coletados das bases de dados SciELO (*Scientific Electronic Library Online*) e DEDALUS (Banco de Dados Bibliográficos da USP). Assim, a partir dessa comparação, torna-se possível determinar quais são as variações e possíveis inadequações presentes nos textos produzidos por brasileiros. O objetivo específico deste estudo foi determinar como os médicos brasileiros usaram o verbo *submit*, uma tradução *prima facie* da

palavra “*submeter*”, e detectar quais estruturas gramaticais foram usadas no *corpus* de referência para expressar o mesmo significado, ou seja, como o sentido de tais estruturas foi concretizado nos textos das revistas publicadas no exterior. Optamos por estudar em detalhes um item lexical, conforme explicado na Metodologia. Em uma etapa posterior, buscamos determinar se há aspectos que precisariam ser modificados nos textos publicados em inglês nas revistas médicas do Brasil, para que sejam produzidos textos com conteúdo linguístico mais próximo daqueles dos artigos publicados em inglês na América do Norte e na Europa.

A partir desses objetivos, as questões de pesquisa a serem investigadas são elencadas abaixo:

- i. Quais são os trigramas mais recorrentes nos textos médicos escritos em inglês publicados em revistas científicas do Brasil?
- ii. Quais trigramas desse conjunto de textos não são encontrados nas revistas médicas escritas em inglês, publicadas nos EUA e Europa?
- iii. Que padrões foram usados nos textos publicados no exterior para expressar o sentido contido nos trigramas exclusivos, encontrados somente nos textos publicados no Brasil?

3 Delimitação do campo de pesquisa

Com relação às concepções epistemológicas, o termo Linguística Aplicada (doravante LA) possui diferentes interpretações. Nesta pesquisa, entendemos LA, a princípio, como Área Interdisciplinar. Kaplan (1980), conforme citado por Celani (1992) faz uma analogia da LA com o teatro: da mesma forma que todas as artes se encontram e se tornam realidade no palco, a LA é considerada como o centro dos estudos da linguagem, o ponto onde a LA se intersecciona com as outras disciplinas. Assim, a interdisciplinaridade está ligada ao surgimento de um problema que, para ser solucionado, atrai para junto de si várias outras disciplinas, buscando a solução a partir dos conhecimentos das diversas disciplinas.

Além disso, adotamos a definição de Moita Lopes (2006) que entende a LA como “mestiça”, pois o autor considera o movimento que o conduziu à necessidade de pensar em uma LA que dialogasse com teorias que estão atravessando o campo das Ciências Sociais e das Humanidades.

Recorrendo à metáfora de palco teatral citada anteriormente, na qual diferentes disciplinas entram em cena para a produção de uma nova realidade, podemos entender esta pesquisa como pertencendo à área de Linguística Aplicada. (ALAMBERT, inédito) e, como citado, a LA tem natureza interdisciplinar, pois coloca outras disciplinas a serviço da resolução dos problemas relacionados à linguagem. Esta pesquisa está ligada à Linguística de *Corpus* (doravante LC), que é um ramo da LA. Como citado anteriormente, na presente pesquisa foram utilizadas as teorias e ferramentas da Linguística de *Corpus*. A LC “é uma área que se ocupa da exploração de *corpora* (conjuntos de dados linguísticos textuais coletados segundo critérios pré-estabelecidos) para pesquisa de uma ou mais línguas” (Berber Sardinha, 2004). Mais especificamente, o trabalho desenvolvido fundamentou-se na exploração da linguagem através de evidências empíricas, extraídas por meio de computador. (BERBER SARDINHA, 2000a)

O objeto primordial de pesquisa da LC é a linguagem em sua manifestação natural e empírica, compilada e processada através de recursos computacionais. Graças ao advento dos computadores e dos programas gerenciadores de *corpus*, atualmente podemos armazenar grandes volumes de dados linguísticos e, a partir de sua análise segundo critérios pré-estabelecidos, é possível tirar conclusões reproduzíveis a partir de material empírico.

O surgimento da Linguística de *Corpus* é uma evidência de que o progresso nesse campo de atuação relativa à linguagem dá-se em oposição à postura de promover a teoria sempre em primeiro lugar. Owen (1993) afirma que a Linguística de *corpus* se apresenta como uma nova abordagem da descrição gramatical, que penetra nas partes da linguagem inacessíveis a outras gramáticas. Assim, a partir da análise do *corpus* de estudo, comparando-o ao *corpus* de referência, poderemos depreender aspectos talvez ainda não acessíveis por outros métodos.

A atividade de versão no contexto brasileiro

Esta pesquisa, a partir da análise de dois conjuntos de artigos científicos da área médica, redigidos em inglês e publicados em periódicos do exterior e do Brasil, tem como foco estudar como os médicos brasileiros redigem seus textos em uma língua estrangeira (no caso, o inglês). Sendo assim, este trabalho irá lidar com textos traduzidos, mais especificamente, abordando a análise de textos traduzidos, utilizando para tal análise os recursos da Linguística de *Corpus*. Entre os trabalhos importantes na área, podemos citar Baker (1995), Bowker (1998), Berber Sardinha (2000a, 2004), Laviosa (1998). Para Alambert (2008, p.10) “a aproximação entre a Linguística de *Corpus* e os Estudos de Tradução não é apenas teórica, mas envolve aspectos práticos muito bem fundamentados que levam à produção de resultados consistentes e não subjetivos”.

Em língua portuguesa, temos dois termos equivalentes a *translation*, que levam em consideração a direção em que se traduz: ao traduzir de uma língua estrangeira para a língua materna, usamos o termo *tradução* e ao passar um texto em língua materna para uma língua estrangeira, dizemos *versão*. Essa diferenciação é bastante importante, uma vez que certos autores defendem que o tradutor deve fazer apenas traduções, e não fazer versões. Newmark (1998) é um autor que afirma que a única língua para a qual o tradutor estaria habilitado a traduzir seria sua língua materna.

No Brasil, entretanto, a prática de versão é uma atividade aceita tanto nos meios acadêmicos quanto nos profissionais. Prova disso são as disciplinas relacionadas à versão presentes nas grades curriculares de cursos de graduação em Tradução (licenciatura) (UNIP; UNAERP) e também em cursos de especialização em tradução (ANHANGUERA; UNIFRAN) de instituições de ensino superior de São Paulo, capital e interior. Além disso, o Sindicato dos Tradutores (SINTRA, data não especificada), em um artigo que visa analisar quais seriam os aspectos importantes na formação dos tradutores, afirma que existe uma Necessidade de “mais atividades de tradução e versão de textos” para uma melhor formação dos tradutores. A JUCESP, por sua vez, que é o órgão ligado à Junta Comercial

que regulamenta as atividades dos tradutores públicos e intérpretes comerciais (os ditos tradutores juramentados) determina, em sua tabela de honorários, um preço diferenciado para as atividades de tradução e de versão (esta última, com valores mais altos), o que reafirma que a versão não é considerada prática antiética no Brasil.

Em termos de mercado de trabalho, no Brasil, existe uma grande necessidade de versões para o inglês, e tanto tradutores experientes quanto inexperientes estão fazendo esses trabalhos (SERRALVO, 2004). Em nossa experiência prática na área de textos médicos, muitas vezes os profissionais de saúde do Brasil, por estarem habituados a ler artigos científicos em inglês, dispensam a ajuda de um tradutor, e produzem seus próprios textos médicos em inglês. Alguns desses profissionais, por iniciativa própria ou por orientação dos periódicos onde pretendem publicar, utilizam os préstimos profissionais de tradutores para a revisão final de seus textos. Como não existem dados concretos que nos permitam aquilatar ao certo qual a incidência de textos produzidos diretamente pelos médicos, quantos passam por revisões e quantos são enviados para publicação sem terem passado por um revisor de língua inglesa, não é possível analisar essas variáveis (autor do texto científico, autor da versão em inglês, existência de revisão).

4 Organização da dissertação

A redação do texto desta dissertação está dividida nos seguintes capítulos:

CAPÍTULO 1, no qual são expostos em detalhes os fundamentos teóricos que servem de embasamento para o presente estudo, mostrando os principais conceitos e estudos desenvolvidos nas áreas de Linguística de *Corpus* e análise de *clusters*.

No CAPÍTULO 2, são apresentados detalhadamente os procedimentos metodológicos adotados para a coleta dos *corpora* e para a análise dos dados obtidos. No início do capítulo é apresentada uma

descrição minuciosa dos *corpora* de estudo (BRAZIL) e de referência (ABROAD), os procedimentos para a montagem dos *corpora*, citando os problemas enfrentados na definição dos textos que compuseram os *corpora*, e as ferramentas utilizadas em sua preparação e análise. A seguir, detalha os procedimentos de análise dos dados, com a definição de certos aspectos fundamentais à pesquisa, como a definição da extensão dos *clusters* (trigramas, quadrigramas, pentagramas).

O CAPÍTULO 3 descreve e analisa os trigramas mais frequentes e os trigramas exclusivos do *corpus* BRAZIL, para tentar determinar quais seriam os termos ou estruturas no *corpus* ABROAD que corresponderiam a tais trigramas. Desses trigramas exclusivos, foram obtidas as linhas de concordância, para podermos compreender o contexto em que eles foram produzidos. A seguir, é apresentada a análise semântico-gramatical das frases que continham os trigramas exclusivos do *corpus* BRAZIL. Por fim, pesquisamos o *corpus* ABROAD, procurando recuperar nesses textos quais os termos ou estruturas que os autores que publicam no exterior utilizam para expressar as ideias ou sentidos correspondentes aos encontrados no *corpus* BRAZIL.

As CONSIDERAÇÕES FINAIS concluem esta dissertação, apresentando a conclusão dos resultados, retomando alguns pontos relevantes, traçando as metas para pesquisas futuras e fazendo um fechamento do estudo.

Os itens REFERÊNCIAS e APÊNDICES encerram esta dissertação.

CAPÍTULO 1 Fundamentação Teórica

O propósito deste capítulo é descrever os preceitos teóricos que embasam esta pesquisa, estando organizado da seguinte maneira: primeiramente são apresentados os princípios da Linguística de *Corpus* relevantes ao presente estudo e as principais definições de LC e alguns conceitos importantes dessa área do saber. Em um segundo momento é apresentado o programa *WordSmith Tools* e seus principais componentes (ferramentas). A seguir, faz-se uma explanação sobre a análise de *clusters*. Em um momento final, são abordados aspectos ligados à redação de artigos científicos.

1.1 A Linguística de *Corpus*

Inicialmente, explicitaremos a área básica na qual esta pesquisa encontra-se inserida, ou seja, a Linguística de *Corpus* (SINCLAIR, 1991; TOGNINI-BONELLI, 2001). A Linguística de *Corpus* é um campo que se dedica à criação e análise de *corpora* (plural da palavra em latim *corpus*), que são conjuntos de textos e transcrições de fala armazenadas em arquivos de computador (BERBER SARDINHA, 2004). A Linguística de *Corpus* considera como seu objeto primordial de pesquisa a linguagem em sua manifestação autêntica, natural e empírica, compilada e processada através de recursos computacionais.

A Linguística de *Corpus* investiga a linguagem, nos seus mais diversos níveis, disponibilizando grandes quantidades de dados anteriormente inacessíveis, mas, hoje passíveis de serem compilados e organizados, graças à informática e, em especial, a alguns programas criados especificamente para esse fim. Sem todos esses recursos informáticos, a Linguística de *Corpus* como conhecemos na atualidade não poderia existir. Portanto, o linguista de *corpus* depende de programas de computador para tratar os dados compilados em *corpora*, evidenciando sempre que a análise e a interpretação dos dados serão feitas pelo pesquisador, de modo que os programas de computador apenas ajudam a sistematizar a coleta, tratar os dados e processar as informações (retirar

partes indesejadas de cada texto, selecionar parte dos arquivos, gerar linhas de concordância, listas de palavras, palavras-chave, inserir e remover etiquetas, entre outros).

Entretanto, segundo Berber Sardinha (2000a, p. 325), “*havia corpora antes do computador, já que o sentido original da palavra ‘corpus’ é ‘corpo’, ‘conjunto de documentos’ (conforme o dicionário Aurélio)*”. Assim, de acordo com este autor, outros *corpora* foram também produzidos em um passado distante, antes mesmo do advento do computador.

Maia (2002) afirma que o uso de *corpus* pode ser útil para tradutores, pesquisadores, professores de idiomas e linguistas, mas cada um desses profissionais irá usar os *corpora* de maneiras diferentes. No presente estudo, usaremos os *corpora* para analisar as similaridades e diferenças existentes entre o inglês utilizado em periódicos médicos publicados no Brasil e textos de mesmo gênero textual e público-alvo similar, escritos em inglês, publicados em revistas médicas do exterior.

Conrad (1996) mostra como o uso de *corpus* pode ampliar as perspectivas da compreensão sobre a variação linguística – pela pesquisa das interações das características observadas em diversos textos – com o objetivo de discutir as implicações dessa perspectiva para a compreensão da variação textual dos escritos acadêmicos e para uma análise mais aprofundada das práticas pedagógicas. Segundo Conrad, as análises baseadas em *corpus* são pesquisas empíricas sobre o uso da língua e têm certas características em comum: são baseadas em textos autênticos, usam ferramentas computacionais para análise automática e interativa e incluem análises qualitativas e interpretações funcionais para descrever padrões característicos de determinada língua. Além disso, é possível pesquisar padrões característicos de diferentes tipos textuais e compreender a variação em textos de um mesmo campo ou gênero textual.

No caso da pesquisa descrita aqui, os recursos da Linguística de *Corpus* proporcionaram condições para que a linguagem de dois *corpora* seja estudada, tornando possível determinar as semelhanças e as divergências entre os padrões léxico-gramaticais usados em ambos os *corpora*. Um trabalho relevante para esta pesquisa, e que ajudou a traçar os rumos a serem tomados foi, entre outros, aquele compilado para formar o

livro Linguística de *Corpus*, de Tony Berber Sardinha (2004), que aborda importantes temas da LC e analisa em detalhes o programa *WordSmith Tools*, versão 3.0, que foi usado para o gerenciamento dos *corpora* coletados. Além de uma visão histórica da área (fundamental para a compreensão da evolução dos conhecimentos relacionados), o livro fornece definições e classificações extremamente úteis para esta pesquisa. A especificidade e a adequação dos *corpora* coletados são dois aspectos que devem ser abordados antes do início da coleta dos textos, para garantirem que esta recolha seja feita segundo os critérios pré-estabelecidos e, assim, o produto final seja representativo e produza resultados confiáveis.

Outros trabalhos utilizaram a LC para a análise de textos da área da saúde. Shergue (2003) utiliza a Linguística de *Corpus* para analisar textos médicos em inglês. Seu enfoque volta-se para a competência linguística específica no inglês como forma de transpor a dificuldade de compreensão oral em apresentações em congressos. Enfocando a Análise do Discurso, Vieira (2002) estuda uma situação de trabalho de médicos infectologistas com pacientes soropositivos e portadores de AIDS, a fim de retomar o processo de produção de sentido a partir da análise enunciativo-discursiva das práticas languageiras dessa população de falantes e Otranto (2006) procura caracterizar o gênero de atividade do dentista e suas relações estabelecidas a partir dos discursos, utilizando aspectos relacionados à co-construção de sentidos na interação profissional.

Embora esses trabalhos citados tenham analisado textos (orais ou escritos) produzidos por profissionais de saúde, não encontramos pesquisas que procurem determinar, com base em LC, as características presentes nos textos médicos escritos em inglês publicados no Brasil, que enfoquem os aspectos que poderiam contribuir para o aperfeiçoamento linguístico dos médicos brasileiros ao redigirem em inglês. O presente estudo busca preencher essa lacuna.

1.1.1 Terminologia relacionada à Linguística de *Corpus*

Como já citado anteriormente, este estudo está baseado nos pressupostos teóricos da Linguística de *Corpus*, além de lançar mão de

alguns de seus recursos metodológicos. Entretanto, como o assunto dos textos coletados é a Medicina e, portanto, o público que possa se interessar em ler esta dissertação, além de linguistas, também inclui os profissionais de saúde ou aqueles que redigem textos para eles (tradutores, revisores, jornalistas, entre outros), consideramos que seja importante fazermos uma pequena apresentação dos principais termos e expressões da área, que foram utilizados durante esta dissertação.

A LC, como qualquer outra área do saber, utiliza uma série de termos cujas definições serão apresentadas a seguir, pois sua compreensão é importante para a leitura deste e dos capítulos posteriores e para o entendimento do presente estudo, seus resultados e conclusões. O objetivo desta breve lista de termos não é aprofundar-se na terminologia da LC, mas sim servir de orientação para leitores que não dominem os termos relacionados a ela, como médicos e estudantes de medicina, que possam se interessar por este estudo. Como em qualquer área do conhecimento, na LC também há divergências no que se refere à definição de termos, e muitos deles são usados em outras áreas da Linguística, na Filosofia, na Psicologia, com sentidos diferentes ou levemente modificados. Assim, este glossário sucinto procura delimitar certos conceitos, localizando-os para o campo da LC, além de determinar um grupo estável de definições, que serão utilizadas em todos os capítulos deste trabalho. São apresentados resumidamente alguns termos de Linguística de *Corpus* utilizados nesta dissertação, com uma breve definição e alguns exemplos elucidativos.

a. *Corpus*

Entre os conceitos mais importantes em LC aqui citados está o de *corpus*. Segundo Berber Sardinha (2004), *corpus* é um conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Assim, nota-se a importância do estabelecimento de critérios prévios à coleta dos textos que irão compor os *corpora*. Tais critérios devem ser baseados nos propósitos da pesquisa que se pretenda fazer, de modo a criarmos uma base de dados

que possa responder nossas questões de pesquisa, sem ser tendenciosa, mas que tenha sido formada tendo em vista tais objetivos.

Para Crystal (1991, p. 86), *corpus* é uma coletânea de “dados linguísticos, sejam textos escritos ou uma transcrição de falas gravadas, que pode ser utilizada como ponto de partida para a descrição linguística ou como um meio de verificar hipóteses sobre uma língua” (*tradução nossa*).

A partir desta segunda definição, muito poderia ser dito, mas iremos nos ater aos pontos mais relevantes para o presente estudo:

- *collection*: se *corpus* é uma coletânea, conforme a definição, isto significa que sua formação é feita a partir de princípios e regras, ou seja, montar um *corpus* não é simplesmente “juntar um monte de palavras”, e “quantas mais, melhor”. A coleta dos textos deve seguir princípios predeterminados, para que possamos, em uma etapa posterior, conseguir chegar a conclusões a respeito dos textos do *corpus* e, o mais importante, para que possamos extrapolar essas conclusões para outros textos que se enquadrem nos mesmos princípios e que compartilhem características semelhantes aos dos textos do *corpus* de estudo.

- *starting-point*: o *corpus* é apenas um ponto de partida, e não o estudo pronto. Nada será obtido do *corpus* se seu conteúdo não for analisado, estudado e ponderado. O *corpus* sem o pesquisador e suas reflexões não passa de um amontoado de palavras sem utilidade.

- *linguistic description*: o estudo de *corpus* tem caráter descritivo, jamais prescritivo. Um *corpus* mostra o que é, e não o que deveria ser (PERROTTI-GARCIA, 2007), deixando as prescrições para as gramáticas tradicionais e para os livros didáticos.

b. Linguística de *Corpus* (LC)

Segundo Berber Sardinha, a Linguística de Corpus

se ocupa da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio do computador.

(BERBER SARDINHA, 2000b, p. 46)

Assim, a partir dessa afirmação inicial, ficam implícitos os seguintes pré-requisitos básicos para que um conjunto de textos seja considerado como objeto de estudo da LC:

Textos autênticos, fruto de uma seleção cujos critérios de coleta tenham sido definidos previamente e com textos representativos da linguagem, de uma variedade linguística ou de um idioma (KENNEDY, 1998).

c. Palavra (ou termo) de busca

É a palavra (ou termo) digitada na janela do programa em que se está trabalhando (no caso deste estudo, o *WordSmith*) e que estará presente, centralizada, nas linhas de concordância. Quando a palavra de busca encontra-se na linha de concordância, é denominada **nódulo**.

d. Linhas de concordância

As linhas de concordância são as ocorrências de um item específico e seu contexto, ou seja, linhas formadas pela palavra de busca, centralizada, e por um número predeterminado de palavras à esquerda e à direita desta. A ferramenta *Concord*, do programa *WordSmith* por exemplo, fornece seus resultados na forma de linhas de concordância, ou seja, linhas de texto contendo a palavra (ou termo) de busca (denominada nódulo) centralizada, com um determinado número de palavras à direita e à esquerda. Precisamos

explicitar dois conceitos usados na LC e nesta pesquisa para a análise das concordâncias: horizonte e cotexto. As palavras ao redor do nóculo que serão levadas em conta como **colocados** do nóculo formam um espaço que, na Linguística de *Corpus*, chamamos de **horizonte**. Por exemplo, nas linhas de concordância abaixo, observamos o nóculo *intend*, com cinco palavras à direita e à esquerda, formando um horizonte de 11 palavras.

data at Baylor, and I intend to milk it as much

*Conclusions— This document does not intend to change the guidelines but
smoked were less likely to intend to breastfeed and to initiate*

Por sua vez, **cotexto** é a quantidade de palavras à direita e à esquerda do nóculo, e essa quantidade pode ser aumentada, bastando para isso acionar um comando do programa *WordSmith Tools*. Não é necessário restringir-se à largura visível da concordância, pois ela pode, quase sempre, ser alargada na tela do computador. Entretanto, isso não costuma ser necessário para nossa análise, mesmo porque Sinclair, no OSTI Report publicado em 1970, (SINCLAIR, 2004) já havia demonstrado que um horizonte de quatro ou cinco palavras de cada lado do nóculo é suficiente para dar conta da padronização da maioria das palavras.

e. Tokens (formas ou vocábulos) e **types** (itens ou ocorrências)

Para a Linguística de *Corpus*, a quantidade total de palavras de um texto ou *corpus* representa o número de *tokens*. Cada vocábulo é contado uma vez, mesmo que haja outros iguais. Por outro lado, cada palavra (diferente) usada em um texto é um *type* (SINCLAIR, 1991). Assim na contagem dos *types*, as palavras iguais somente são contadas uma vez. Vejamos por exemplo a estrofe do poema *Sacred Emily*, de Gertrude Stein:

Rose is a rose is a rose is a rose.

Esta estrofe tem dez palavras, mas apenas três formas diferentes: “*rose*”, “*is*” e “*a*”. Assim, teremos 10 *tokens* (ou seja, 10 palavras, independente de serem iguais ou diferentes entre si) e apenas três *types*,

uma vez que o termo *token* corresponde aos itens (ou ocorrências) e *types* às formas (ou vocábulos) (BERBER-SARDINHA, 1999).

Peirce (1932) foi o primeiro autor a utilizar os termos *tokens* (itens) e *types* (formas) com o sentido empregado posteriormente pela LC. Esses termos são originados da Filosofia. Naquele campo, *types* refere-se a sentidos abstratos e únicos, enquanto *tokens* são concretos. (WETZEL, 2007).

f. Lista de frequência de palavras

É uma lista que contém todas as palavras de um determinado *corpus*, que pode estar classificada em ordem de frequência ou alfabética. Na lista de palavras classificadas por ordem decrescente de frequência, as palavras de maior ocorrência aparecem primeiro, seguidas sucessivamente pelas de menor ocorrência. Na lista de palavras classificadas por ordem alfabética, as palavras iniciadas pela letra A aparecem primeiro (como em um dicionário), seguidas sucessivamente pelas iniciadas pelas letras B, C, D, e assim sucessivamente. Cada vez que a ferramenta *WordList* é acionada – ou seja, ao clicar em *Make a WordList now* - o programa *WordSmith Tools* (WST) fornece duas listas de frequência: (A), lista das palavras classificadas em ordem alfabética; (F), lista das palavras classificadas em ordem (decrescente) de frequência. Além dessas, o programa também gera, simultaneamente, uma terceira lista, (S), que contém diversos dados estatísticos sobre o *corpus* analisado (por exemplo, número de *tokens*, número de *types*, número de sentenças e de parágrafos, e número de palavras com uma letra, duas letras, três letras, até 25 letras ou mais, entre outras informações estatísticas sobre os textos que formam o *corpus* individualmente e em sua totalidade).

g. Clusters, bundles, chunks ou n-gramas

A linguagem tem um caráter probabilístico, o que significa dizer que certas palavras, expressões e sequências de palavras têm uma maior probabilidade de ocorrer em certos tipos de texto do que outras, com uma determinada frequência, e por isso revelam padrões de combinação dessas palavras em uma determinada língua, gênero textual ou variante.

Ao detectarem a presença de “sequências de palavras” que ocorrem em uma frequência elevada em certos tipos de textos, diversos autores procuraram criar termos para nomear essas sequências recorrentes. A seguir analisaremos alguns desses termos e, por fim, trataremos de n-gramas, que foi a denominação adotada neste estudo.

Expressões fixas

Haswell (1991) utilizou o termo “expressões fixas” para denominar as combinações recorrentes de palavras e seu uso recorrente refletiria, segundo o autor, uma certa maturidade na escrita, podendo ser considerado como uma marca de proficiência em determinado idioma.

Prefabs

De Cock *et al.* (1998) denominaram *prefabs* as sequências de palavras facilmente memorizadas, para uso pelo falante com maior velocidade de comunicação. Para os autores daquele estudo, o enfoque foi a maneira como o cérebro humano classificaria e recuperaria as informações linguísticas no momento da produção (oral ou escrita).

Expressões formulaicas

De Cock *et al.* (1998) identificaram o que eles chamaram de “expressões formulaicas”. Após fazer a lista de frequência das “sequências”, os autores fizeram a “filtração” por identificação “manual”, ou seja, foram

fazendo a leitura das expressões para sua classificação. Nessa fase inicial, foram eliminadas as repetições não-intencionais, os fragmentos de frases ou de expressões. Além disso, os autores avaliaram se as expressões remanescentes tinham potencial para exercerem funções pragmáticas ou discursivas. Um aspecto interessante dessa pesquisa foi a detecção de que algumas expressões eram utilizadas excessivamente pelos aprendizes (em relação aos especialistas) enquanto outras eram subutilizadas. Os autores concluíram que, no *corpus* estudado, os aprendizes realmente utilizaram “*prefabs*” e, em alguns casos, até mais do que os escritores experientes, mas que muitas dessas expressões não eram exatamente as mesmas usadas pelos especialistas, não eram usadas na mesma frequência, eram aplicadas com usos sintaticamente diversos ou exerciam funções pragmáticas diferentes.

Lexical bundles

Biber, Conrad & Cortes (2004) definiram os *lexical bundles* como sequências multipalavras que seguem padrões estruturais bastante constantes. Segundo Cortes (2006), Os *lexical bundles* são extremamente frequentes na escrita acadêmica, mas raramente são usados por estudantes universitários. Biber & Conrad (1999, p. 188) consideram que os feixes lexicais são “tão comuns”, que podemos considerar erradamente que sejam expressões simples e, portanto, que serão aprendidos com facilidade. Mas, segundo Cortes (2006) a aquisição dos feixes lexicais não é um processo natural, mesmo sendo altamente frequentes em textos acadêmicos (CORTES, JONES & STOLLER, 2002). Os *lexical bundles* são extremamente frequentes nos *corpora* de artigos publicados em periódicos estadunidenses reconhecidos nas áreas de História e Biologia (CORTES, 2004). Entretanto, seu uso pelos alunos, que eram todos nativos de Inglês, em diferentes níveis universitários, foi extremamente raro.

Clusters* ou *n-gramas

Scott (1997) definiu os *clusters* como sequências de palavras recorrentes em um texto com uma determinada frequência, podendo ter de dois a, no máximo, oito elementos. Segundo Partington (2006), *clusters* são “sequências de palavras (em geral de dois a um máximo de oito itens)” que evidenciam “maneiras típicas de dizer coisas e, portanto, típicas de determinado autor/falante.” Scott & Tribble (2006) explicam como utilizar *clusters* para a pesquisa de contrastes entre os textos produzidos por especialistas e aprendizes, em um *corpus* de textos acadêmicos que incluiu as dissertações de alunos de mestrado em literatura polonesa e artigos de crítica literária do *corpus* BNC.

Após realizarem a análise de um *corpus* de aprendiz, Scott & Tribble (2006) concluíram que a análise de *clusters* em textos comparáveis pode ser útil para a identificação de áreas nas quais os autores desejem melhorar seu desempenho, comparando seus textos com a escrita que aspiram ter. Este achado ajuda a justificar o uso de *corpora* comparáveis na presente pesquisa, uma vez que os autores não fazem a idolatria ao escritor nativo, mas ponderam que, do contato e da análise dos textos feitos por autores experientes, o aprendiz pode se beneficiar e conseguir indícios de como chegar a um padrão almejado.

Com o uso de programas de computador (por exemplo, *WordSmith*), é possível determinar os *n-gramas* (*clusters*) mais recorrentes em *corpora*. *N-gramas* são combinações de palavras, ou seja, consistem em sequências de “n” palavras pertencentes a um texto (CANDIDO Jr., 2007). Tais *n-gramas* recebem o nome correspondente ao número de itens que o compõem, por exemplo, se $n=2$, são chamados de bigramas; se $n=3$, são denominados trigramas; $n=4$, quadrigramas, e assim sucessivamente. O número máximo teórico de itens para um *n-grama* é oito. Os *n-gramas* são extraídos automaticamente pelo programa escolhido, sendo apresentados geralmente em ordem de frequência.

O estudo dos trigramas ajuda a expor os elementos constituintes dos *corpora*, revelando sua estrutura e permitindo uma comparação em um nível mais profundo da estrutura da língua, a partir do estudo da ocorrência e recorrência de determinadas estruturas ou palavras. A análise dos trigramas exclusivos do *corpus* BRAZIL procura evidenciar características idiossincráticas dos textos das revistas médicas publicadas no Brasil, procurando, em uma etapa posterior, determinar no *corpus* de referência quais seriam as estruturas empregadas em substituição a elas.

Por tratar-se de *corpora* com mais de seis milhões de palavras cada, as ferramentas computacionais tiveram um papel muito importante na concretização deste estudo, mas a Linguística de *Corpus* não foi usada apenas como metodologia, e sim como arcabouço teórico para a determinação dos *corpora* de estudo e de referência, para a análise dos trigramas mais frequentes e dos trigramas exclusivos, e também para a determinação dos colocados e sua transposição para o *corpus* de referência. Essa assim chamada retextualização significa determinar, no *corpus* de referência, quais seriam os padrões empregados para expressar o sentido contido nos trigramas exclusivos.

h. Padrões léxico-gramaticais

Em textos de escritores aprendizes, para diferentes autores (BAMBERG, 1983; McCULLY, 1985), o uso de combinações recorrentes de palavras é considerado um sinal de proficiência no uso da língua em determinados registros, entre os quais os gêneros acadêmicos. Mas, segundo Schleppegrell (1996), os alunos não-nativos de inglês, para terem sucesso em contextos acadêmicos, precisam dominar os padrões léxico-gramaticais presentes nos textos dos gêneros acadêmicos.

A LC baseia-se em uma visão da linguagem como sistema probabilístico (JOHNS, 1986 *apud* BERBER SARDINHA, 2004), ou seja, embora muitas ocorrências possam ser possíveis, algumas têm maior probabilidade de ocorrer. O aspecto de maior relevância da diferença de frequência de ocorrência é o fato de a variação não ocorrer aleatoriamente.

Ou seja, a linguagem se organiza segundo padrões léxico gramaticais que ocorrem em determinadas frequências, dependendo do tipo de texto, gênero, registro e, muito importante no presente estudo, nível de domínio do idioma. (BERBER SARDINHA, 2004)

A padronização tem como base a probabilidade de ocorrência das palavras. Assim, estudar os padrões léxico-gramaticais significa analisar do ponto de vista estatístico quais são as escolhas mais frequentes em um determinado tipo textual.

CAPÍTULO 2 Metodologia

O intuito deste capítulo é apresentar os procedimentos metodológicos utilizados para a coleta de textos e análise dos *corpora*, descrevendo-os em detalhes. Primeiramente serão apresentados os objetivos da pesquisa e como foi feita a escolha do campo a ser pesquisado. A seguir, tratamos das ferramentas utilizadas e dos procedimentos metodológicos para a coleta do *corpora* e para a análise dos dados obtidos.

A metodologia adotada é *corpus-driven*, o que significa que o estudo não partiu de um inventário preexistente de características ditas problemáticas na escrita dos autores ou de trigramas previamente selecionados, mas que lançamos mão de recursos da Linguística de *Corpus* para determinar, a partir de textos autênticos, comparados com textos de referência, tidos com um padrão de prestígio, quais seriam os itens lexicais problemáticos no *corpus* de estudo e quais seriam as escolhas feitas no *corpus* de referência para exprimir os mesmos sentidos.

2.1 Objetivos da pesquisa

Conforme já mencionado na Introdução, a pesquisa descrita aqui teve como objetivo geral identificar os padrões léxico-gramaticais característicos de textos médicos escritos em inglês publicados em revistas do Brasil e, a seguir, procurar determinar se tais padrões seriam recorrentes em textos de mesmo tipo textual (artigos científicos), registro (aqui entendido como o nível de formalidade do texto) e público alvo (profissionais de saúde) similar, publicados nos Estados Unidos e Europa. Além disso, o presente estudo procurou analisar aspectos exclusivos dos textos médicos publicados no Brasil e determinar quais seriam as estruturas ou termos correspondentes, nos textos publicados em revistas médicas do exterior.

Além de fornecer o arcabouço teórico que embasou este estudo, os princípios da Linguística de *Corpus* (LC) também são utilizados para o

tratamento dos dados, sua análise e interpretação. A LC, segundo Hunston (2002), propicia uma abordagem empírica de observação de dados, em formato eletrônico, que poderão ser analisados tanto quantitativamente quanto qualitativamente, permitindo que seja feito um levantamento de padrões léxico-gramaticais da língua estudada, bem como a determinação dos padrões recorrentes ou exclusivos de cada *corpora*.

Acreditamos que os trabalhos científicos publicados em inglês em revistas editadas em países anglófonos possam ser considerados um ponto de partida para a comparação, a partir do qual é possível determinar a distância/proximidade entre os textos publicados em revistas médicas brasileiras e os publicados em revistas internacionais, do ponto de vista da padronização léxico-gramatical. Empregaremos este conceito de 'distância', como uma metáfora que significa "a diferença entre a padronização exibida nos textos das revistas internacionais e aquela exibida nos textos das revistas nacionais", e vice-versa. Assim, o objetivo específico foi fazer uma análise das diferenças entre ambos os *corpora* para, posteriormente, sugerir quais seriam os aspectos que poderiam ser modificados nos textos escritos em inglês nas revistas médicas publicadas no Brasil, para que possam ser produzidos textos mais próximos dos publicados em inglês na América do Norte e na Europa, considerados pela comunidade médica como um modelo de prestígio.

2.2 Escolha da área de pesquisa

A área de Medicina foi escolhida para ser o foco desta pesquisa primeiramente por ser um campo em que atuo há mais de duas décadas, tanto como profissional da saúde, em uma primeira etapa, quanto como tradutora e professora de ESP (*English for Specific Purposes*) e EAP (*English for Academic Purposes*), de modo que temos uma bagagem bastante relevante, que nos permite lidar com esse assunto com segurança e bastante julgamento crítico. Além disso, é cada vez maior o número de revistas científicas brasileiras nas áreas de Medicina e Odontologia, que optam por apresentar seu conteúdo em língua inglesa, o que torna

necessária uma investigação da redação desses textos, sob vários ângulos, no nosso caso, do ponto de vista da padronização léxico-gramatical, a fim de termos um parâmetro da proximidade idiomática entre o texto de periódicos brasileiros frente a um padrão de prestígio internacional.

2.3 Seleção dos *corpora*

A pesquisa baseada em *corpus* fornece evidências concretas sobre as características recorrentes nos textos que compõem os *corpora* analisados. Para que os *corpora* sejam representativos, a coleta dos textos deve ser baseada em critérios estabelecidos antes do início da coleta. Desde que os textos de cada *corpus* sejam coletados de modo a representar da melhor maneira possível os fenômenos linguísticos mais significativos e recorrentes no tipo textual sob análise, o *corpus* final, fruto do processo de coleta, também será representativo e, assim, será útil para os propósitos pretendidos.

A coleta de textos para compor os dois *corpora* utilizados neste estudo (*corpus* de estudo, denominado *corpus* BRAZIL, e *corpus* de referência, denominado *corpus* ABROAD) obedeceu a rígidos critérios de seleção, que serão detalhados a seguir. É importante enfatizar que tais critérios foram estabelecidos antes do início da coleta dos textos, para que todos os artigos científicos coletados estivessem de acordo com tais critérios, de modo a dispormos de dois *corpora* que pudessem estar de acordo com os requisitos da LC para sua produção.

2.3.1 Critérios para a montagem dos *corpora*

a. Critérios de inclusão

Conforme explicitado no capítulo anterior, uma pesquisa baseada em LC fundamenta-se no uso de *corpora* que preencham objetivos estabelecidos previamente à sua coleta. Ambos os *corpora* utilizados no presente trabalho contêm textos com as seguintes características:

- especializados (artigos científicos da área médica), redigidos por médicos, ou a seu pedido, e tendo como público-alvo pretendido os profissionais da área da saúde;
- comparáveis (ambos os *corpora* compartilham uma série de características similares: tamanho, público-alvo, gênero textual, assunto dos artigos científicos); um dos *corpora* (BRAZIL) foi composto por textos publicados em revistas médicas brasileiras, sendo considerado como o *corpus* de estudo, e o outro (ABROAD) por revistas médicas publicadas nos EUA e Europa, sendo este último considerado como *corpus* de referência;
- monolíngues (só foram coletados textos escritos em inglês, mesmo quando fossem procedentes de revistas bilíngues). Entretanto, para que houvesse uma quantidade similar de artigos em inglês britânico e americano, o *corpus* de textos publicados em revistas da América do Norte e da Europa (ABROAD) foi balanceado, de modo a conter aproximadamente 50% de textos redigidos em inglês britânico e 50% em inglês norte-americano.

Além disso, foram coletados apenas artigos completos (as referências foram eliminadas, por considerarmos que estas, não tendo sido escritas originalmente pelos autores dos artigos, não refletiriam a linguagem usada por eles), textos indexados nas bases de dados PubMed e DOAJ, para o *corpus* ABROAD; e SciELO e DEDALUS para o *corpus* BRAZIL.

Com relação à data de publicação, incluímos artigos publicados após 2006. O caráter diacrônico dos textos que integram ambos os *corpora* é inerente, segundo Stubbs (2004), pois a rigor os dados são registros de atos de comunicação que já ocorreram, mesmo que há muito pouco tempo. Entretanto, a coleta procurou contemplar textos publicados apenas a partir de 2005, para refletir a escrita contemporânea tanto das revistas publicadas em inglês no Brasil quanto no exterior.

Os textos de ambos os *corpora* (BRAZIL, *corpus* de estudo e ABROAD, *corpus* de referência) foram selecionados especialmente com a finalidade de servirem de base para esta pesquisa.

No presente estudo, levamos em consideração o local de publicação das revistas científicas, ao invés da nacionalidade dos autores dos textos publicados nas referidas revistas. Assim, para efeitos desta pesquisa, os textos publicados nas revistas brasileiras, desde que escritos em inglês, serão parte do *corpus* BRAZIL, enquanto os textos publicados nas revistas médicas dos EUA e Europa serão incluídas no *corpus* ABROAD, independentemente do país de origem e da nacionalidade dos autores de cada artigo.

b. Critérios de exclusão

Não foram coletados:

- artigos de revistas publicadas antes de 2006;
- artigos incompletos;
- artigos de periódicos não indexados nas bases de dados PubMed e DOAJ (*corpus* ABROAD) e SciELO e DEDALUS (*corpus* BRAZIL);
- artigos de revistas científicas das áreas de Veterinária e Medicina Animal.

2.3.2 Perfil dos *corpora* de estudo e de referência

Seguindo os critérios pré-estabelecidos citados anteriormente, foram incluídos textos de 14 revistas científicas publicadas no Brasil (totalizando 5.075.197 palavras) e de 31 revistas científicas publicadas no exterior (EUA e Europa) (sendo 2.679.374 palavras nas revistas escritas em inglês britânico e 2.613.366 palavras em inglês norte-americano, totalizando 5.292.740 palavras). (Tabelas 1, 2 e 3).

REVISTAS BRASILEIRAS	TOKENS
Revista Brasileira de Anestesiologia	177.668
<i>São Paulo Medical Journal</i>	348.776
Jornal de Pediatria	584.519
<i>Brazilian Dental Journal</i>	260.917
Jornal Vascular Brasileiro	161.944
Revista Brasileira de Medicina do Esporte	285.542
<i>Brazilian Journal of Infectious Diseases</i>	493.757
Revista do Instituto de Medicina Tropical de São Paulo	25.161
<i>Brazilian Journal of Medical and Biological Research</i>	1.287.600
Revista Brasileira de Psiquiatria	213.833
<i>International Brazilian Journal of Urology</i>	254.702
Arquivos Brasileiros de Endocrinologia e Metabologia	225.189
Revista Brasileira de ORL	284.082
<i>Clinics</i>	471.507
TOTAL DE TOKENS	5.075.197

Tabela 1: Revistas científicas que compuseram o *corpus* de estudo (BRAZIL, com revistas brasileiras). A coluna à direita mostra o número de palavras (*tokens*) de cada revista.

REVISTAS EM AMERICAN ENGLISH	TOKENS
<i>American Journal of Clinical Oncology</i>	288.502
<i>Stroke</i>	670.153
<i>Pediatrics (American Academy of Pediatrics)</i>	709.320
<i>Clinical Diabetes</i>	119.924
<i>American Journal of Cardiology</i>	267.768
<i>Clinical Medicine & Research</i>	114.584
<i>The Internet Journal of Emergency Medicine</i>	32.686
<i>The Internet Journal of Third World Medicine</i>	78.283
<i>American Journal of Nephrology</i>	297.191
<i>American Journal of Otolaryngology: Head Neck Medicine and Surgery</i>	34.955
TOTAL DE TOKENS	2.613.366

Tabela 2: Revistas científicas que compuseram o *subcorpus* de referência com revistas médicas publicadas no exterior, com artigos em inglês norte-americano [*American English*]. A coluna à direita mostra o número de palavras (*tokens*) de cada revista.

REVISTAS EM <i>BRITISH ENGLISH</i>	TOKENS
BMC	155.132
British Medical Journal	171.236
Head and Face Medicine	214.124
AIDS Research and Therapy	189.298
Journal of Vascular Research	225.045
Journal of Neuroinflammation	253.501
BioPsychoSocial Medicine-	70.542
International Journal of Behavioral Nutrition and Physical Activity	386.646
Geriatrics	23.300
Endocrine Disorders	19.780
Nuclear Medicine (UK)	35.732
The Lancet	195.365
British Medical Journal of Ophthalmology	6.584
British Journal of Ophthalmology	284.302
Emergency Medicine	23.912
Immunity and ageing	40.854
Pediatric Rheumatology	79.768
Journal of Occupational Medicine and Toxicology	58.927
Cough	24.595
Blood Disorders	21.532
Orphanet Journal of Rare Diseases	199.199
TOTAL DE TOKENS	2.679.374

Tabela 3: Revistas científicas que compuseram o *subcorpus* de referência com revistas médicas publicadas no exterior, com artigos em inglês britânico [*British English*]. A coluna à direita mostra o número de palavras (*tokens*) de cada revista.

Houve a preocupação de coletarmos textos variados, quanto aos assuntos, para que houvesse artigos representativos das diferentes áreas da medicina humana em ambos os *corpora*. A Tabela 4 apresenta um resumo dos temas abordados pelas revistas científicas de ambos os *corpora*. O número entre parênteses significa a quantidade de volumes incluída para cada periódico. É importante lembrar que a quantidade de *tokens* de ambos os *corpora* foi bastante próxima, e foi o critério de término de coleta mais importante. O número de volumes (e conseqüentemente o número de artigos) foi maior no *corpus* BRAZIL, pois os artigos eram menos extensos. Sendo assim, precisamos de mais volumes para obter um número de *tokens* similar ao do *corpus* ABROAD.

Foram incluídos textos de Odontologia no *corpus* BRAZIL, para contrabalançar os textos de Medicina Bucal encontrados nas revistas dos EUA.

<i>Corpus</i> BRAZIL	<i>Corpus</i> ABROAD
Anestesiologia (6)	Cardiologia (16)
Medicina Geral (47)	Medicina Geral (31)
Pediatria (14)	Pediatria (9)
Odontologia (7)	Oncologia (4)
Vascular (6)	Vascular (6)
Urologia (5)	Nefrologia (6)
Medicina do Esporte (7)	Medicina Ocupacional e dos Esportes (3)
Moléstias Infecciosas (12)	AIDS (2)
Medicina Tropical (1)	Medicina do “Terceiro Mundo” (5)
Psiquiatria (6)	Geriatria (1)
Endocrinologia (1)	Endocrinologia (1)
Otorrinolaringologia (5)	Otorrinolaringologia (3)
	Oftalmologia (6)
TOTAL: 12 temas (117 volumes)	TOTAL: 13 temas (94 volumes)

Tabela 4: Temas abordados pelas revistas científicas das quais foram coletados os textos para compor os *corpora*. O número entre parênteses significa a quantidade de volumes incluída para cada periódico.

A coleta de ambos os *corpora* (BRAZIL e ABROAD) levou cerca de oito meses (totalizando 246 horas de coleta, em 57 períodos de três a cinco horas cada), e levou em consideração critérios determinados previamente, conforme citado. Perrotti-Garcia (2005), ao abordar o uso de *corpora* customizados como fonte de pesquisa para tradutores, reafirma a importância da determinação prévia dos critérios de coleta de textos para comporem o *corpus* de estudo.

2.3.3 Problemas enfrentados na definição dos textos que compuseram os *corpora*

Em uma etapa preliminar, ainda durante a definição do projeto de pesquisa, a ideia era utilizar apenas resumos de artigos científicos (*abstracts*), que são facilmente encontrados na *web*, gratuitamente e em

grande quantidade. Entretanto, após uma análise da terminologia e da fraseologia desses textos, notamos que essa abordagem não corresponderia às expectativas, sendo insuficiente para atender às nossas necessidades. Assim, optamos por coletar artigos científicos completos (*full papers*), mesmo sabendo que essa coleta seria bem mais complexa e trabalhosa.

Um aspecto importante que deve ser enfatizado é a relativa escassez de textos técnicos de alto nível (de especialista para especialista) e de textos acadêmicos de certas áreas do saber que sejam de acesso livre pela Internet – o que dificulta muito a coleta de material para a criação de *corpora* com essas características (MAIA, 2002). Principalmente levando-se em consideração que, como já dissemos anteriormente, os textos deveriam ser completos (muitos periódicos disponibilizados tiveram que ser descartados, pois as bases de dados apenas liberavam parte do conteúdo do artigo – em geral apenas os Casos Clínicos ou o capítulo de Material e Métodos).

Além disso, outro problema enfrentado foi relacionado ao caráter sincrônico dos textos coletados, ou seja, apenas foram coletados textos publicados a partir de 2005, o que tornou ainda mais difícil coletar textos completos, de acesso livre, pela Internet (em geral, alguns periódicos mais consagrados, apenas disponibilizam textos de volumes publicados há mais de cinco ou dez anos, fazendo com que a busca por textos mais recentes seja bastante complexa).

Outra preocupação durante a coleta dos textos foi balancear o conteúdo do *corpus* de estudo e do *corpus* de referência, de modo a conterem um número aproximado de artigos das diferentes especialidades da área da saúde. Além disso, como em alguns países estrangeiros a Odontologia é uma subespecialidade da Medicina (a assim chamada Medicina Bucal), também foram incluídos textos dessa área.

2.4 Ferramentas computacionais utilizadas na preparação e análise dos dados

O presente estudo utiliza *corpora* comparáveis. Com o uso do programa *WordSmith*, é possível determinar os n-gramas (*clusters*) mais recorrentes nos *corpora* de estudo e referência (BERBER SARDINHA, 1999). N-gramas, como citado anteriormente, são combinações de palavras, ou seja, consistem em sequências de “n” palavras pertencentes a um texto (CANDIDO Jr., 2007). Os n-gramas foram extraídos automaticamente, pelo programa *WordSmith*, dos *corpora* de estudo e referência, sendo apresentados em ordem de frequência. No presente estudo, utilizamos trigramas, ou seja, *clusters* com três elementos. A seguir, utilizando uma programação em linguagem Shell, foram determinados os trigramas exclusivos de cada *corpora*, e, em uma etapa posterior, com outra programação, foram recuperadas as frequências de ocorrência desses trigramas na listagem original do *WordSmith*. Na fase seguinte, agora manualmente, os trigramas exclusivos foram analisados, procurando determinar padrões de ocorrência de cada trigrama. Esses padrões foram, então, pesquisados no *corpus* de referência (ABROAD), para determinarmos quais seriam as opções dos autores dos textos publicados nas revistas internacionais para criarem frases que tenham o significado pretendido pelos autores dos trigramas exclusivos do *corpus* BRAZIL.

2.4.1 O programa *WordSmith Tools*

Dentre os diversos *softwares* disponíveis para auxiliar o linguista de *corpus*, um deles se destaca: *WordSmith Tools*. Primeiramente, por ser um conjunto de ferramentas, utilitários e acessórios muito úteis; em segundo lugar, por ter sido idealizado especificamente para trabalhar com *corpus*, segundo a concepção terminológica da LC; e também por ser um programa bastante difundido, o que facilitaria a reprodução de um estudo por outros pesquisadores, que contariam com os mesmos recursos, sem precisar

lançar mão de muitos programas diferentes para conseguir ter à sua disposição o conjunto de ferramentas contidas nesse programa.

O programa, desenvolvido por Mike Scott, da Universidade de Liverpool, no Reino Unido, apresenta, entre outras, três ferramentas: *WordList*, *KeyWords* e *Concord*. As ferramentas a serem utilizadas no presente estudo são *WordList* e *Concord*. A primeira delas produz três listas: lista das palavras em ordem alfabética (A); lista das palavras em ordem de frequência (decrecente) (F); lista que contém diversos dados estatísticos sobre o *corpus* analisado (S).

A partir da escolha de uma palavra (ou termo) de busca, a ferramenta *Concord* fornece seus resultados na forma de linhas de concordância, com o nóculo centralizado e a possibilidade de escolher a extensão do horizonte (total de palavras à esquerda e à direita do nóculo). Além disso, também existe a possibilidade de reclassificar os resultados por ordem alfabética, levando em consideração as palavras à direita ou à esquerda do nóculo, em uma sequência predeterminada pelo pesquisador (ou seja, por exemplo, 1L2D, classificar primeiramente pela primeira palavra à esquerda e a seguir pela segunda à direita do nóculo, e assim sucessivamente).

Com relação ao uso de linhas de concordância, Stubbs (2004) reforça seu uso em uma ampla gama de áreas; contudo, o autor reafirma que não há descobertas automáticas, mas as generalizações indutivas podem ser testadas contra as observações procedentes dos *corpora* coletados. Portanto, o presente estudo parte da análise de trigramas, situa os trigramas em seu contexto (linhas de concordância) e utiliza estas linhas como dados para a posterior análise e classificação dos padrões encontrados.

No Quadro 1, reproduzido e adaptado de Berber Sardinha (1999, p. 89), encontram-se listadas as ferramentas, utilitários e instrumentos disponíveis no *WordSmith Tools*, bem como suas funções.

Ferramentas, utilitários, instrumentos e funções disponíveis no *WordSmith Tools*

Componentes: O *WordSmith Tools* é composto por (a) ferramentas, (b) utilitários, (c) instrumentos, e (d) funções.

Há três **ferramentas** e quatro **utilitários**, nomeadamente:

(a) **Ferramentas:**

WordList ; KeyWords; Concord

(b) **Utilitários:**

Renamer; Text Converter; Splitter; Viewer

(c) **Instrumentos** de análise disponíveis (com os nomes em inglês entre parênteses):

..*WordList:*

1. Lista de palavras individuais ('*wordlist*').
2. Lista de multipalavras ('*wordlist, clusters activated*').
3. Lista de palavras de consistência individuais ('*detailed consistency*').
4. Lista de multipalavras de consistência ('*detailed consistency, clusters activated*').
5. Lista de dimensões e densidade lexical ('*statistics*').

..*Concord:*

1. Concordância ('*concordance*').
2. Lista de colocados ('*collocates*').
3. Lista de agrupamentos lexicais ('*clusters*').
4. Lista de padrões de colocados ('*patterns*').
5. Gráfico de distribuição da palavra de busca ('*plot*').

..*KeyWords:*

1. Lista de palavras-chave ('*keywords*').
2. Banco de dados de listas de palavras-chave ('*database*').
3. Lista de palavras-chave chave ('*key keywords*').
4. Lista de palavras-chave associadas ('*associates*').
5. Lista de agrupamentos textuais ('*clumps*').
6. Gráfico de distribuição de palavras-chave ('*keyword plot*').
7. Listagem de elos entre palavras-chave ('*keyword plot links*').

As principais **funções** (d) distribuídas nas três ferramentas são:

1. Lematização: agrupamento de duas ou mais formas diferentes em um mesmo item.

2. Classificação: ordenação de listas e concordâncias

3. Delimitação: escolha de quais partes do *corpus* serão lidas pelo programa.

Quadro 1: Ferramentas, utilitários, instrumentos e funções disponíveis no *WordSmith Tools*, reproduzido de Berber-Sardinha (1999, p. 89), com adaptações, grifos nossos.

O programa *WordSmith Tools*, desenvolvido por Mike Scott da Universidade de Liverpool, foi utilizado para a obtenção dos *clusters* (trigramas). A seguir, como precisávamos obter a lista dos trigramas exclusivos do *corpus* de estudo, e o WordSmith não disponibiliza esse recurso, desenvolvemos rotinas escritas na linguagem de programação Shell, que foram rodadas no programa Cygwin (que pode ser usado em ambiente Windows, como se estivéssemos utilizando o sistema operacional Unix). A linguagem Shell possibilitou a criação de um *script* para determinarmos os trigramas exclusivos do *corpus* BRAZIL. O *script* desenvolvido é apresentado no Apêndice C.

2.5 Definição da extensão dos *clusters*

Um dos critérios definidos para a análise do *corpus* foi a extensão dos *clusters*. A extensão de um *cluster* refere-se ao número de termos que irão compô-lo. Scott & Tribble (2006) analisaram os *clusters* na escrita acadêmica em inglês. Fizeram uma análise das 20 palavras de maior ocorrência em dois *corpora* e notaram que ambas eram muito similares. Os autores fizeram a mesma análise para *clusters* de duas, três e quatro palavras (bigramas, trigramas e quadrigamas, respectivamente). Os autores concluíram que o mais indicado é analisar trigramas (em uma análise dos colocados à direita, para estudar o contraste entre os diferentes estilos de escrita ou o produto dos diferentes grupos de escritores) e quadrigamas (que são fortes fatores discriminantes entre diferentes registros).

Com base nos achados citados, optamos por trabalhar apenas com os trigramas, uma vez que todos os artigos científicos incluídos nos *corpora* foram redigidos por médicos e para médicos, e não faremos uma análise do registro (ou seja, do nível de formalidade) dos textos e nossa abordagem está relacionada aos diferentes estilos de escrita, ou o produto dos diferentes grupos de escritores.

2.6 Procedimentos metodológicos gerais da pesquisa

Optamos por trabalhar com os trigramas mais frequentes em ambos os *corpora*, como ponto de partida, e a seguir com os trigramas exclusivos do *corpus* BRAZIL, pois seriam aqueles menos típicos, menos usuais e inesperados – onde talvez possamos encontrar evidência léxico-gramatical que indique a distância entre o texto de brasileiros e o texto de revistas de prestígio. Com relação ao uso dos trigramas mais frequentes, Stubbs (2004) afirma que a repetição de certas ocorrências é um achado significativo na análise dos *corpora*. A análise quantitativa irá revelar o que é típico, usual e esperado.

Após a determinação dos trigramas de ambos os *corpora* pelo programa *WordSmith*, e dos trigramas encontrados exclusivamente no *corpus* BRAZIL, utilizando-se rotinas em linguagem Shell e o programa *Cygwin*, os trigramas exclusivos do *corpus* BRAZIL foram analisados manualmente. Desses trigramas exclusivos, foram obtidas as linhas de concordância, para podermos compreender seu cotexto e recuperarmos o contexto em que foram produzidos. Desse modo, a seguir, obtivemos as linhas de concordância que continham os trigramas exclusivos do *corpus* BRAZIL. Nessa etapa, verificamos a léxico-gramática dos textos e optamos pelo estudo detalhado de um item lexical, que era bastante frequente no *corpus* BRAZIL, com uma frequência bastante baixa no *corpus* ABROAD.

A partir da determinação do item lexical que seria analisado em profundidade, fizemos um levantamento manual dos colocados desse item. Notamos que haviam relações entre os colocados de cada nóculo, ou seja, os trigramas analisados estabeleciam relações com determinadas palavras, talvez apontando para um problema de interferência da língua materna no momento da produção das frases em inglês. Em uma etapa posterior, de posse dos colocados, pesquisamos o *corpus* ABROAD, procurando recuperar nesses textos quais seriam os termos ou estruturas utilizadas nos textos publicados no exterior para expressar os sentidos correspondentes aos encontrados no *corpus* BRAZIL. Ou seja, o processo foi o de tentar, a

partir dos colocados dos trigramas exclusivos do *corpus* BRAZIL, determinar quais seriam os termos ou estruturas correspondentes no *corpus* ABROAD. Assim, procuramos determinar quais eram os padrões léxico-gramaticais presentes nos textos publicados nas revistas internacionais que corresponderiam ao que os autores das revistas brasileiras haviam feito de modo "exclusivo".

2.7 Convenções adotadas para a apresentação dos padrões

Os padrões léxico-gramaticais presentes nos textos publicados nas revistas internacionais são apresentados no capítulo seguinte. Para uma melhor visualização dos padrões, utilizamos alguns recursos gráficos e de formatação. Vejamos, por exemplo, no padrão a seguir

patient(s) + TAKe/en/ing + (complementos) + *therapy

- os parênteses [()] significam que as palavras que estão dentro deles podem ou não estar presentes simultaneamente;
- símbolo de adição [+] significa soma, ou seja, o padrão é formado tanto pelos elementos que antecedem o sinal quanto pelos que o sucedem;
- as barras [/] significam que o padrão pode ser formado pelos elementos que a antecedem ou pelos que a sucedem;
- a palavra “complementos” [complementos], quando presente no padrão, significa que podem haver outros elementos (substantivos, numerais, pronomes ou adjetivos) situados nas posições indicadas;
- o asterisco [*] colocado antes de uma palavra indica que esta pode ser precedida por um prefixo, e colocado após uma palavra significa que esta pode apresentar sufixos;
- as palavras escritas com letras maiúsculas [TAK] indicam radicais verbais, que para formarem o verbo flexionado necessitam de morfemas verbais complementares.

CAPÍTULO 3 Análise de Dados

Este capítulo mostra a análise e a interpretação dos dados necessária para responder as perguntas de pesquisa. Em um estudo-piloto (PERROTTI-GARCIA, 2008) foi avaliado um *subcorpus* composto por 61.151 *tokens* procedentes de 24 textos coletados de duas revistas da área médica (*Sao Paulo Medical Journal* e *Brazilian Journal of Infectious Diseases*), ambas publicadas no Brasil, para determinarmos a validade da utilização da classificação funcional de Biber (BIBER, 2003) no presente estudo. A classificação de Biber foi aplicada a textos acadêmicos orais e escritos (transcrição de aulas, conversações, livros de texto e prosa acadêmica), razão pela qual foi a escolhida como ponto de partida para a análise, uma vez que os textos também são do campo acadêmico. Entretanto, a partir do referido estudo piloto, pudemos observar que tal classificação não seria suficiente para a análise dos trigramas de nossos *corpora*, razão pela qual optamos por uma análise mais voltada para a significação de determinados trigramas exclusivos do *corpus* de estudo, sem procedermos à sua classificação prévia.

Para a análise dos dados obtidos, inicialmente lançamos mão de diversas tabelas, procurando resumir de modo mais facilmente visualizável os resultados das análises preliminares. Por tratar-se de *corpora* muito extensos, mesmo as etapas que foram feitas de modo manual (pela leitura e análise individualizada de cada trigrama ou linha de concordância) tiveram de contar com recursos de computador, com tabelas feitas no programa Excel e, em uma etapa posterior, com gráficos para procurar ilustrar melhor os resultados obtidos.

A seguir, são discutidos os resultados da aplicação do procedimento de análise, considerados à luz dos princípios teóricos que embasam este estudo, procurando responder às perguntas de pesquisa feitas na Introdução, considerando os resultados frente aos objetivos.

3.1 Trigramas mais frequentes no *corpus* de referência

O *corpus* ABROAD, contendo textos médicos de revistas científicas publicadas no exterior foi considerado, pelas razões expostas anteriormente, como o *corpus* de referência neste estudo.

Com o uso da ferramenta *WordList*, do programa *WordSmith Tools*, com a opção *clusters* ativada e selecionada a extensão de três itens (n=3), foram obtidos 5.292.738 trigramas, dos quais o Apêndice D apresenta os 85 trigramas mais frequentes, que ocorreram em uma frequência igual ou superior a 0,01% para referência do leitor, devido à impossibilidade de imprimir todo o conteúdo de todas as listas usadas no estudo. Cabe salientar que, para a análise subsequente, dos *clusters* exclusivos, foram levados em conta todos os *clusters* retirados pelo programa *WordSmith Tools*, conforme o critério explicitado no capítulo de metodologia. Note-se que, por um aspecto ligado à concepção do programa *WordSmith*, alguns numerais encontrados nos textos processados pelo programa são convertidos no símbolo # (jogo da velha ou cerquilha), como pode ser observado, por exemplo, no trigrama *at least #* do referido Apêndice. Nesse caso, especificamente, o símbolo # representa o número 12, o trigrama original era *at least 12*, e o numeral foi substituído. Entretanto, o programa não substituiu todos os números, apenas alguns, e aparentemente faz isso de modo aleatório.

3.2 Trigramas mais frequentes no *corpus* de estudo

O *corpus* BRAZIL, contendo textos médicos de 14 revistas científicas publicadas no Brasil, (totalizando 5.075.197 palavras) foi considerado, pelas razões expostas anteriormente, como o *corpus* de estudo. Com o uso da ferramenta *WordList*, do WST, com a opção *clusters* ativada e a extensão de três itens selecionada, foram obtidos 5.075.195 trigramas, dos quais o Apêndice E apresenta uma amostra para referência apenas, contendo os 111 trigramas mais frequentes, que ocorreram em uma frequência igual ou superior a 0,01%.

3.3 Trigramas exclusivos do *corpus* BRAZIL

Ao utilizarmos a ferramenta *WordList*, do WST, com a opção *clusters* ativada e a extensão de 3 itens selecionada, foram obtidos mais de cinco milhões de trigramas no *corpus* BRAZIL, entretanto, como nosso objetivo era determinar quais eram as características idiossincráticas deste *corpus*, utilizando as ferramentas computacionais descritas no capítulo de metodologia, obtivemos uma listagem dos trigramas exclusivos do referido *corpus*, que totalizaram 2.503 trigramas. Entendemos por trigramas exclusivos do *corpus* de estudo aqueles que ocorreram somente no *corpus* BRAZIL. Dessa listagem, o Apêndice F apresenta os 185 trigramas exclusivos do *corpus* BRAZIL encontrados com uma ocorrência igual ou superior a 100, que serão nossa principal fonte de material para análise a partir de então.

O Apêndice G apresenta os primeiros 25 resultados obtidos, ao determinarmos os trigramas exclusivos do *corpus* BRAZIL, utilizando a programação em Shell e o programa Cygwin, já citados anteriormente. Como é possível observar no apêndice citado, o arquivo resultante, embora trouxesse os trigramas exclusivos do *corpus* BRAZIL, os apresentava em ordem alfabética, e sem as informações referentes ao número de ocorrências de cada trigrama e sem sua frequência relativa no *corpus* BRAZIL. Por essa razão, para recuperar esses dados, na etapa seguinte, foi preciso desenvolver outro *script* em Shell que recuperasse essas informações. Para tanto, contamos com a colaboração do meu orientador, o Prof. Dr. Tony Berber Sardinha, que desenvolveu uma rotina, em linguagem Shell, para que pudéssemos recuperar os dados citados. O Apêndice H apresenta o *script* desenvolvido por meu orientador, o Prof. Tony, para a recuperação das frequências dos trigramas exclusivos do *corpus* BRAZIL.

Após utilizar a programação em linguagem Shell para recuperar as frequências de ocorrência dos trigramas exclusivos do *corpus* BRAZIL, obtivemos um arquivo contendo os trigramas apresentados em ordem decrescente de frequência e suas frequências de ocorrência.

A Tabela 5 apresenta os primeiros 25 resultados obtidos, com o número de ocorrências de cada trigrama.

Trigramas exclusivos do <i>corpus</i> BRAZIL (número de ocorrências): primeiros 25 resultados
LARGER VERSION OF (1540)
K JPG FILE (1531)
JPG FILE (1446)
THIS IMAGE # (974)
IMAGE # K (973)
DE SÃO PAULO (818)
OF SÃO PAULO (814)
RIO DE JANEIRO (573)
OF THIS TABLE (553)
THIS TABLE # (537)
TABLE # K (537)
FILE TABLE (514)
CORRESPONDENCE ABSTRACT (502)
UNIVERSITY OF SÃO (500)
WERE SUBMITTED TO (460)
UNIVERSIDADE FEDERAL DE (445)
FACULDADE DE MEDICINA (425)
ADDRESS FOR CORRESPONDENCE (421)
CORRESPONDENCE AND FOOTNOTES (418)
ARE REPORTED AS (417)
AND FOOTNOTES (417)
FOOTNOTES ABSTRACT (414)
UNIVERSIDADE DE SÃO (408)
FILE FIGURE (399)
FEDERAL UNIVERSITY OF (396)

Tabela 5: Trigramas exclusivos do *corpus* BRAZIL; em ordem decrescente de ocorrência e com o número de ocorrências entre parênteses, primeiros 25 resultados obtidos.

3.4 Análise dos trigramas contendo o verbo *submit*

Ao analisar os 111 trigramas de maior ocorrência no arquivo Só Brazil (ou seja, o arquivo que continha os trigramas exclusivos do *corpus* BRAZIL), observamos que cinco deles (perfazendo 1.111 ocorrências) continham o verbo *submit*, no particípio (*submitted*). Com finalidade comparativa, levantamos as ocorrências de *submit* no *corpus* ABROAD. Foi utilizado o caractere coringa (*) na busca feita com a ferramenta *Concord* do programa *WordSmith*, de modo a todas as ocorrências de *submit* (mesmo no particípio) serem recuperadas.

A seguir são apresentados os cinco trigramas encontrados apenas no *corpus* BRAZIL que continham o verbo *submit*, em qualquer conjugação.

Trigramas exclusivos do *corpus* BRAZIL com o termo *submit** (*número de ocorrências*):

WERE SUBMITTED TO (460)

SUBMITTED TO THE (168)

PATIENTS SUBMITTED TO (166)

WAS SUBMITTED TO (162)

SUBMITTED TO A (155)

Diante desses achados e em consonância com a metodologia *corpus-driven* seguida neste estudo, optamos por analisar cada um dos trigramas do *corpus* BRAZIL que continham o verbo *submit*. Tal análise é apresentada a seguir.

3.4.1 Análise do trigrama *were submitted to*

Ao fazermos a análise de todas as ocorrências do trigrama *were submitted to*, tomando por base as primeiras 25 linhas de concordância obtidas no *corpus* BRAZIL podemos observar os achados apresentados no

Quadro 2, onde a primeira coluna apresenta os agentes da locução verbal passiva *were submitted to* e a terceira coluna o objeto indireto.

QUEM	TRIGRAMA	A QUE
patients; individuals; children; cases; adolescents	WERE SUBMITTED TO	administrations of; a serological test to; preoperative chemotherapy schedule; examination; hysterectomy; temporal lobectomy; treadmill gait training; a new procedure; amputation ; complete blood count; the same preoperative chemotherapy schedule; preoperative chemotherapy; Doppler ultrasound examination; distal revascularization-interval ligation; cross- spectral analysis
animals, rabbits; hens;		euthanasia; a procedure; conventional treatment
Haemophilus spp		susceptibility test

Quadro 2. Padrões apresentados no *corpus* BRAZIL pela análise de todas as ocorrências do trigrama *were submitted to* nas primeiras 25 linhas de concordância (a primeira coluna apresenta os agentes da locução verbal passiva *were submitted to* e a terceira coluna o objeto indireto).

Nota-se que, na primeira coluna (QUEM), existe referência a seres humanos, animais e microorganismos, enquanto na terceira coluna (A QUE), existem diversos procedimentos e exames. Desse modo, fica evidenciado que o verbo *submit* está sendo usado com uma tradução direta de “submeter-se” (como em “o paciente foi submetido a uma cirurgia”).

3.4.2 Análise do trigrama *was submitted to*

Ao fazermos a análise de cada um dos trigramas em separado, em relação ao trigrama *was submitted to*, tomando por base as primeiras 25 linhas de concordância obtidas no *corpus* BRAZIL podemos observar os achados apresentados no Quadro 3, onde a primeira coluna apresenta os agentes da locução verbal passiva *was submitted to* e a terceira coluna o objeto indireto.

QUEM	TRIGRAMA	A QUE
<i>only the patient from case ; the LP group the remainder</i>	WAS SUBMITTED TO	<i>surgery; DS, temporary tracheotomy; urgent endivascular treatment; adrenalectomy; myocardial revascularization; laparoscopic left radical nephrectomy</i>
<i>the alloy; DNA; protein; cDNA; total saline extract antigen</i>		<i>previous oxidation; a second round of purification; reverse-phase HPLC; the polymerase chain reaction; electrophoresis</i>
<i>Each animal; Another group of animals</i>		<i>stimulations; only one stimulation; the same protocol</i>
<i>data set collected; data</i>		<i>acoustic analysis; two-way ANOVA</i>

Quadro 3. Padrões apresentados no *corpus* BRAZIL pela análise de todas as ocorrências do trigrama *was submitted to* nas primeiras 25 linhas de concordância (a primeira coluna apresenta os agentes da locução verbal passiva *were submitted to* e a terceira coluna o objeto indireto).

Nota-se que, na primeira coluna (QUEM), assim como já foi observado na análise do trigrama *were submitted to*, existe referência a seres humanos, animais e microorganismos, além de elementos da bioquímica (proteína, DNA, extrato de antígeno), enquanto na terceira coluna (A QUE), existem diversos procedimentos e exames, além de algumas provas estatísticas. Desse modo, fica evidenciado que o verbo *submit*, também nos trigramas de *was submitted to*, está sendo usado com uma tradução direta de “submeter-se” (como em “o paciente foi submetido a uma cirurgia”).

3.4.3 Análise dos demais trigramas contendo a palavra *submit**

Com base nas linhas de concordância do trigrama *patients submitted to*, procurando determinar quais seriam as relações estabelecidas entre o sujeito do verbo *submit** (no caso, *patients*) e seus objetos, no *corpus* BRAZIL, construímos o Quadro 4.

	A QUE
<i>PATIENTS SUBMITTED TO</i>	<i>coronary catheterization; video endoscopic inguinal lymphadenoplasty; percutaneous coronary intervention; surgery; arterial revascularization surgery; CEA; immediate endovascular treatment; radical cystectomy; this procedure; tympanoplasties; empirical radioiodine therapy; grafts; surgical procedures; CAS with EPD; 6,006 surgeries; preoperative chemotherapy; carotid endarterectomy; various surgical procedures; ruptured AAA surgery; single Epleys maneuver; TSS ; iliac artery aneurysm</i>

Quadro 4. Padrões apresentados no *corpus* BRAZIL pela análise de todas as ocorrências do trigramma *patients submitted to* nas primeiras 25 linhas de concordância.

Pela análise dos complementos do trigramma, ficou evidenciado que os autores que publicaram artigos nas revistas do Brasil incluídas no *corpus* BRAZIL utilizaram o verbo *submit* com o sentido de submeter-se a um procedimento cirúrgico, diagnóstico ou terapêutico. Houve apenas uma ocorrência na qual o trigramma apareceu associado a uma condição patológica (*iliac artery aneurysm*, ou seja, aneurisma da artéria ilíaca), e mesmo nesse caso o verbo *submit* foi usado com o sentido de “estar sujeito à”.

Ao analisarmos os outros dois trigramas exclusivos do *corpus* BRAZIL que continham a palavra *submit**, observamos que os exemplos seguem os mesmos padrões citados anteriormente, ou seja, o verbo *submit* está sendo usado com uma tradução direta de “submeter-se” (como em “o paciente foi submetido a uma cirurgia”).

3.4.4 Ocorrências de *submit no *corpus* ABROAD**

Passamos, então, a avaliar os dois trigramas que continham o verbo *submit* no *corpus* ABROAD. Para essa avaliação, partimos das linhas de concordância. Foram encontradas quatro ocorrências da palavra *submit* no

corpus ABROAD (duas de cada trigramma). Apresentamos a seguir linhas de concordância em questão.

AUTHORS TO SUBMIT 2 ocorrências

Contextualização:

*You can encourage potential **authors to submit** their manuscripts to*

*Your support is also needed through encouraging potential **authors to submit** their highest quality research*

COMPLETE AND SUBMIT 2 ocorrências de uma mesma frase

Contextualização:

Specifically, all members of the writing group are required to **complete and submit** a Disclosure Questionnaire showing all such relationships that migh...

Observamos que, embora iguais, as frases contendo o trigramma *complete and submit* encontram-se em textos diferentes, de volumes diferentes, de um mesmo periódico.

Como podemos observar pela análise dos dois trigramas apresentados acima, em nenhum deles o verbo *submit* foi usado na voz passiva, sendo que nas quatro ocorrências o verbo tinha o sentido de “enviar (um texto) para análise” ou “enviar para publicação”, enquanto no *corpus* BRAZIL o termo foi utilizado com o sentido de “encaminhar”, “submeter-se” ou “distribuir”. Além disso, no segundo caso (*complete and submit*), as duas ocorrências foram de uma mesma frase, repetida em volumes diferentes da mesma revista. Também chama a atenção o fato de serem apenas quatro ocorrências, em um *corpus* tão extenso, o que reafirma a noção de que esse verbo não aparece no *corpus* ABROAD com o sentido dado pelos autores que publicaram nas revistas brasileiras.

3.5 Retextualização: padrões equivalentes a *submitted*

Uma vez que o verbo *submit* (uma tradução *prima facie* da palavra “*submeter*”), não foi encontrado no *corpus* ABROAD com o sentido atribuído a ele pelos autores que publicaram nas revistas médicas do Brasil, podemos inferir que essa utilização não é uma escolha adequada. Procuramos determinar, então, nos textos do *corpus* ABROAD, de que modo foi expressa

a ideia de que alguém era submetido a um determinado procedimento ou exame, a fim de detectarmos quais padrões léxico-gramaticais foram usadas no *corpus* ABROAD para expressar o mesmo significado.

Afinal, se em ambos os *corpora* os artigos falam de pessoas que são submetidas a exames, procedimentos ou testes, qual (quais) verbo(s), ou estrutura(s) sintática(s), serão encontrados nos textos do *corpus* de referência sendo empregados para expressar uma ideia similar? Para tentar determinar a resposta (ou respostas) a essa questão, decidimos não partir de um dicionário, nossa intuição ou consulta a especialistas, mas sim do próprio *corpus*, novamente em consonância com a metodologia *corpus-driven* adotada nesta pesquisa. Assim, optamos por pesquisar, no *corpus* ABROAD, quais seriam os verbos utilizados para “unir” os elementos apresentados nas colunas “quem” e “a que” dos Quadros 2 e 3. Para determinar de que modo a ideia de “ser submetido” era expressa no *corpus* ABROAD, utilizando a ferramenta *Concord*, pesquisamos algumas das palavras da coluna um (QUEM) simultaneamente à ocorrência das palavras da coluna três (A QUE) em um horizonte de cinco palavras à direita (5R). Para facilitar a detecção dos padrões empregados, ajustamos o programa de modo que os resultados fossem classificados em ordem alfabética, pela primeira e segunda palavras à direita do nóculo (1R, 2R). Assim, as linhas de concordância foram classificadas em ordem alfabética, pelas palavras situadas após o nóculo, nas posições 1 e 2.

Além da palavra de busca, ajustamos a ferramenta *Concord* (na guia *Concordance Settings*) para que fossem pesquisadas palavras que ocorrem junto com o termo de busca (*Context Word[s]*). Por exemplo, em uma primeira busca, o programa pesquisou ao mesmo tempo a palavras *patient(s)* e, em um horizonte de cinco palavras à esquerda ou à direita, a palavra *procedure(s)*; depois *patient(s)* e, em um horizonte de cinco palavras à esquerda ou à direita, a palavra de busca.

Além disso, para que pudéssemos encontrar um maior número de ocorrências, após analisarmos as colunas um e três dos Quadros 2 e 3,

decidimos utilizar o caracter coringa (*), para a busca das palavras no singular e no plural, e também para que encontrássemos um maior número de ocorrências de procedimentos (**ectomy*), tratamentos (**therapy*) e exames (**scopy*) simultaneamente. Também utilizamos os termos *procedure* e *test*, que ocorreram na terceira coluna dos referidos quadros e na do Quadro 4. Assim, utilizando a ferramenta *Concord* do programa WST, foram feitas as pesquisas resumidas no Quadro 5.

<i>Palavra de busca</i>	<i>Palavra de contexto</i>	<i>Total de ocorrências</i>
<i>patient*</i>	<i>*therapy</i>	838
<i>patient*</i>	<i>*ectomy</i>	218
<i>patient*</i>	<i>procedure*</i>	175
<i>patient*</i>	<i>examination</i>	143
<i>patient*</i>	<i>*scopy</i>	28
<i>patient*</i>	<i>test</i>	2

Quadro 5. *Corpus* ABROAD: parâmetros de pesquisa das linhas de concordância (palavras de busca [coluna à esquerda], de contexto [coluna central] e número total de ocorrências [coluna à direita])

O primeiro par de palavras pesquisado foi *patient(s)* e **therapy*. A seguir, como muitos dos procedimentos encontrados na coluna A QUE dos Quadros 2 e 3 continham o sufixo **ectomy*, fizemos uma pesquisa utilizando o caracter coringa (*) e este sufixo, para observarmos qual seria sua ocorrência entre as cinco palavras à direita e à esquerda do nóculo. Também pesquisamos *patient** e *procedure**; depois, o par de palavras *patient** e *examination**. Por fim, pesquisamos o sufixo **scopy* e a palavra *test* entre as cinco posições à direita e à esquerda de *patient**.

A seguir, faremos uma análise detalhada de cada uma das pesquisas cujos parâmetros estão resumidos no Quadro 5, apresentando inicialmente em destaque os dados de cada uma das linhas do quadro, e a seguir os dados de análise, com as linhas de concordância respectivas, para exemplificar os padrões encontrados.

3.5.1 Padrões relacionados à coocorrência de *patient** com **therapy*

Ao pesquisarmos a frequência de uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD, foram detectadas 838 ocorrências. Dessas, selecionamos manualmente as que correspondiam ao significado atribuído a *submit* no *corpus* BRAZIL, ou seja, à ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame).

A seguir, cada um desses verbos ou estruturas será analisado separadamente:

a. Padrões contendo o verbo *receive*

O verbo *receive* foi o mais frequente ao pesquisarmos o uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas 92 ocorrências, distribuídas em oito padrões diferentes. Embora os padrões sejam formados por tempos verbais diferentes, em todos os casos observados as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa, em diversos casos, pelo verbo *submit*.

Padrão: *patient(s)* + *receive* + (complementos) + **therapy*

Exemplos:

1. *patient and the medical staff; 6) patients receive group therapy for social skill trainin*
2. *work of Martin et al [5]) as many patients receive postoperative radiation therapy tha*

Padrão: *patient(s)* + *receiving* + (complementos) + **therapy*

Exemplos:

1. *arterial events (2.1%v 1.0%) in patients receiving chemotherapy and Avastin versus*
2. *ever, the occurrence of IRU in patients receiving anti-CMV therapy draws its use int*

3. *ts placed on surveillance to 70 patients **receiving** adjuvant radiation **therapy**. First, th*
4. *roved survival in hemodialysis patients **receiving** injectable vitamin D **therapy** [42]. T*
5. *consisted of 294 consecutive patients **receiving** long-term statin **therapy** who under*
6. *have also been evaluated in patients **receiving** high dose radiation **therapy** for lung*
7. *tive study, the charts of 107 patients **receiving** post-RRP salvage radio**therapy** were*
8. *was preceded by the patient **receiving** high (myeloablative) doses of chemo**therapy***
9. *in the ipsilateral thorax of patients **receiving** postoperative radio**therapy** was statist*

Padrão: *patient(s) + having received + (complementos) + *therapy*

Exemplo:

1. *breast cancer patients. Two occurred in patients **having received** radio**therapy***

Padrão: *patient(s) + was/were receiving/using + (complementos)
+ *therapy*

Exemplos:

1. *ography). At the time of the study, 18 patients **were using** anti-platelet **therapy***
2. *30 mg daily dose. The remaining 2 patients **were not receiving** paricalcitol **therapy***

Padrão: *patient(s) + received + (complementos) + *therapy*

Exemplos:

1. *survived for 1 year. Both patients **received** 2 months of **therapy** and were then tak*
2. *thnicity. Overall, 13% of patients **received** chemo**therapy** and radiation therapy, 3*
3. *unacceptable toxicity. Neither patient **received** further **therapy**. There was nothing u*
4. *year) [44, 45]. Most of these patients **received** ACE inhibitor **therapy** and those achi*
5. *trexed. The majority of patients also **received** either induction chemo**therapy** bef*
6. *noted in all patients. All patients **received** perioperative β -blocker **therapy**. At our*
7. *therapeutic program All patients **received** cognitive behavioral **therapy** with a beh*
8. *ength determination. All patients **received** neoadjuvant chemoradiation **therapy** at*
9. *ecurrence at 5 years, 3 patients **did not receive** systemic **therapy** and 2 received*

10. **received** radiation **therapy** only. Surprisingly, 37% of all patients **did not receive** either **chemotherapy** or radiation

Padrão: patient(s) + has/have received + (complementos) + *therapy

Exemplo:

1. *ctable NSCLC. Many such patients **have received** preoperative chemoradio**therapy***

Padrão: patient(s) + had received + (complementos) + *therapy

Exemplos:

1. *sites. Forty-eight patients **had received** adjuvant chemotherapy previously.*

2. *the cervix.²⁴ Most of these patients **had received** prior radiation **therapy**. However, i*

Um aspecto interessante foi a utilização da estrutura *patient(s) who*, formando padrões com o verbo *receive*, que nos textos do *corpus* ABROAD apenas foram observados com o verbo no passado. Entretanto, não é possível chegar a conclusões a respeito de uma eventual influência do tempo verbal sobre o uso da estrutura assinalada. A seguir, são apresentadas linhas de concordância nas quais foi observado o uso do referido padrão.

Padrão: patient(s) who received + (complementos) + *therapy

Exemplos:

1. *our series for a better prognosis for patients **who received** immuno**therapy**, but this*

2. *Furthermore, there may have been patients **who received** this **therapy** but did not h*

3. *significant given the proportion of patients **who received** postoperative radio**therapy***

4. *median follow-up of 7 years for the patients **who received** adjuvant radiation **therapy***

5. *the first 4 cycles of therapy. Patients **who received** less than one cycle of **therapy***

Padrão: *patient(s) who had received + (complementos) + *therapy*

Exemplos:

1. *mesothelioma arising in patients who **had received radiotherapy**, the latent perio*
2. *required after surgery. Patients who **had received chemotherapy** or **radiotherapy***

b. Padrões contendo o verbo *treat*

O verbo *treat* foi o segundo mais frequente ao pesquisarmos o uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas 52 ocorrências, distribuídas em três padrões diferentes. Entretanto, os padrões foram formados apenas pelo verbo na voz passiva ou na forma de particípio, exercendo a função sintática de complemento nominal, e sempre com a preposição *with*.

Padrão: *patient(s) + being treated with + (complementos) + *therapy*

Exemplo:

1. *udies performed in patients **being treated with thrombolytic therapy** have*

Padrão: *patient(s) + was/were + treated with + (complementos) + *therapy*

Exemplos:

1. *Postlumpectomy patients **were treated with radiotherapy** just to the ipsilateral br*
2. *are presented. The patients **were treated with photodynamic therapy** for cho*
3. *red with those patients who **were pretreated** with 2 or more palliative ch*

Padrão: *patient(s) + treated with + (complementos) + *therapy*

Exemplos:

1. *their treatment. Most of these patients **treated with radiotherapy** are able to tolerate*
2. *reas the other studies assessed patients **treated exclusively with radiotherapy**. Pre-*
3. *nd to increase the proportion of patients **treated with thrombolytic therapy** to >10%.*
4. *he intra-arterial group. Although patients **treated with intra-arterial therapy** showed*
5. *and biochemical control in patients **treated with permanent prostate brachytherapy**.*
6. *ce of hypothyroidism among patients **treated with external beam radiotherapy** alone*

A linha apresentada a seguir, é um exemplo em que a preposição *with* foi utilizada para expressar a ideia de que os pacientes foram “tratados com” quimioterapia, desse modo, optamos por classificar esta frase como sendo um exemplo do uso do padrão *patient(s) + treated with + (complementos) + *therapy*, no qual houve a elipse do verbo *treat*.

1. *Hypothyroidism occurred in 40% of the **patients with chemotherapy** and in 30% of those*

Mesmo quando a ideia a ser expressa era a de que os pacientes não haviam sido submetidos a determinado tratamento, a estrutura *treated with* foi empregada, com algumas alterações. Como é possível observar nos exemplos a seguir, a palavra *not* foi empregada antes de *treated with*, no primeiro exemplo apresentado a seguir, e o prefixo *un-* foi usado no exemplo seguinte. Já o último exemplo desta sequência mostra o uso de outro prefixo (*pre-*), para expressar a ideia de anterioridade cronológica.

Exemplos:

1. *received radiotherapy and 4 in patients **not treated with radiotherapy**. There were*
2. *DS Patient Eligibility Patients previously **untreated with chemotherapy**, immu*
3. *chemotherapy-naive patients and patients **pretreated with fluorouracil-based chemotherapy***

Nota-se, ainda, na linha de concordância 3 apresentada neste item, uma segunda estrutura (sublinhada), que merece atenção. Observamos que, em uma mesma frase, há dois modos diferentes de expressar a noção de “submeter-se” a tratamento ou terapia, a adjetivação (em *chemotherapy-naive patients*) e o uso do verbo *treat*, com o prefixo *pre-*, e a preposição *with*.

c. Padrões contendo o verbo *undergo*

O verbo *undergo* foi o terceiro mais frequente ao pesquisarmos o uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas 40 ocorrências, distribuídas em cinco padrões diferentes. Embora os padrões sejam formados por tempos verbais diferentes, em todos os casos observados as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s) + undergoing + (complementos) + *therapy*

Exemplos:

1. *Sixty-one case histories of patients **undergoing therapy** for malignant brain tumors*
2. *a role in the protection of patients **undergoing chemotherapy** with neurotoxic agen*
3. *in healthy subjects and in patients **undergoing anticoagulant therapy**. This observa*
4. *Plasma VEGF levels in patients **undergoing ABT-510 therapy**. Each line represe*
5. *permitted CT simulation of patients **undergoing breast conservation therapy** with diff*

Padrão: *patient(s) + underwent + (complementos) + *therapy*

Exemplos:

1. *and December 2002. All patients **underwent brachytherapy** more than 3 years pri*
2. *formed.48 Sixteen of these patients **underwent salvage radiotherapy**, 15 had RIS up*
3. *node found on CT. The patient **underwent** 6 cycles of **chemotherapy** with cyclop*

Outro aspecto interessante foi a utilização da estrutura *patient(s) who*, formando padrões com o verbo *undergo*, em diferentes tempos verbais.

Padrão: *patient(s) who underwent* + (complementos) + **therapy*

Exemplo:

1. *A retrospective survey of patients who **underwent** chemotherapy and radiation for*

Padrão: *patient(s) who has/have undergone* +

(complementos) + **therapy*

Exemplo:

1. *Andi's liver function values. Patients who **have undergone** chemotherapy may*

Padrão: *patient(s) who had undergone* + (complementos) +

**therapy*

Exemplos:

1. *of 1,581 breast cancer patients who **had undergone** chemotherapy at the M. D.*
2. *status of 0, 1, 2, or 3. Patients who **had undergone** adjuvant chemotherapy wit*

Outro exemplo interessante é apresentado na linha a seguir. Nela, observamos o uso do auxiliar *did* com função enfática. Neste exemplo, embora o verbo *undergo* esteja no *Simple Past*, em uma frase afirmativa, foi acrescentado o *did* para expressar a noção de que os pacientes que efetivamente foram submetidos ao procedimento cirúrgico tiveram uma evolução significativamente diferente daqueles que não foram operados.

1. *hemotherapy (12 patients). Patients **who did undergo** surgery showed a significantly*

d. Padrões contendo preposições

O uso de preposições foi uma estratégia frequentemente observada ao pesquisarmos o uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas sete ocorrências, distribuídas em cinco padrões diferentes. Embora não sejam numericamente muitas ocorrências, em todos os casos as frases analisadas preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*. A seguir são apresentados exemplos dessas estruturas.

Padrão: *patient(s)* + *after* + (complementos) + **therapy*

Exemplos:

1. *to image patients **after therapy**, timing to image patients after FDG injection, therapy*
2. *serum and CSF was observed in MS patients **after immunomodulatory therapy** [15].*

O uso da preposição *after* referindo-se à **therapy* cria uma ideia de anterioridade do procedimento terapêutico (terapia) em relação a outros elementos da frase. Desse modo, é possível reproduzir a ideia de que o paciente foi submetido a um determinado procedimento após a realização de um tratamento. No primeiro exemplo, os pacientes foram submetidos a exames de imagem após a terapia; no segundo, os parâmetros laboratoriais foram observados após a terapia com imunomoduladores, ou seja, após o paciente ser submetido a tal terapia.

Outra preposição bastante utilizada foi *on*, expressando a ideia de “sob”, tanto precedida do verbo *to be*, em diferentes tempos verbais, quanto isoladamente. Pela análise realizada, avaliando cada uma das ocorrências, todas expressaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), e sua alta frequência revela que seu uso é bastante idiomático no *corpus* de referência (*corpus* ABROAD).

Padrão: *patient(s) + on + (complementos) + *therapy*

Exemplos:

1. *three therapeutic regimes used patients **on Therapy** I which is the association of a*
2. *only ($P < 0.0001, 0.0001$) in HIV patients **on antiviral therapy** compared to those not*
3. *C that were isolated from uremic patients **on CAPD therapy**. Materials*
4. *trials.8 In their meta-analysis, patients **on cisplatin-based chemotherapy** frequently*
5. *cerebral hemorrhages occurred in patients **on aspirin plus warfarin therapy**; 1*

Padrão: *patient(s) + being on + (complementos) + *therapy*

Exemplo:

1. *The consequence of this is patients **being on antiviral therapy** when the HIV infectio*

Padrão: *patient(s) was/were + on + (complementos) + *therapy*

Exemplos:

1. *er MDT by 336 pg/ml while the patient **was on nifedipine therapy**, but the increase w*
2. *llected and analyzed. None of the patients **were on immunomodulating therapy** or im*

A estrutura *patient(s) who*, formando padrões com o verbo *undergo*, em diferentes tempos verbais também foi observada.

Padrão: *patient(s) who is/are + on + (complementos) + *therapy*

Exemplos:

1. *of acute abdomen/abdominal pain in patients who **are on warfarin therapy**, even*

e. Emprego do verbo *take*

O uso do verbo *take* foi observado em algumas linhas de concordância, ao pesquisarmos o uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda

da palavra *patient(s)* no *corpus* ABROAD. Embora não sejam muitas ocorrências, e em algumas delas o verbo *take* esteja em um tempo verbal diferente das demais, em todos os casos as frases analisadas preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s)* + *TAKe/en/ing* + (complementos) + **therapy*

Exemplos:

1. *pectrum of metabolic changes seen in patients **taking** antiretroviral **therapy** [29].* The
2. *r graft fistulae were excluded. No patient **has taken** estrogen replacement **therapy** or*
3. *can not be certain that the patients **were taking** their **therapy** continuously or correc*
4. *ription of controller to patients who **were** not currently **taking** a controller **therapy***

Notamos, pela análise das linhas de concordância apresentadas, que três frases apresentam o verbo *take* no gerúndio (*taking*), em dois deles sendo usado o tempo *past progressive*, também conhecido como *past continuous*, que expressa uma ação que estava ocorrendo no passado. É interessante notar que a opção pelo verbo *take* parece advir do uso de *therapy* na acepção de medicamento, remédio ou fármaco.

f. Emprego do verbo *have*

O verbo *have*, como verbo principal, foi observado em algumas linhas de concordância, ao pesquisarmos o uso da palavra *therapy* (e dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Em todos os casos analisados, o verbo *have* estava no *Simple Past* e as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*. Em alguns casos, foi acrescentada a palavra *who*, formando a estrutura *patients who* como sujeito do verbo *have*.

Padrão: *patient(s) (who) + had + (complementos) + *therapy*

1. *... to 12 patients. None of the patients **had** maintenance chemotherapy. Median*
2. *... B-04 protocol were there patients who **had** radiotherapy following a total mastecto*
3. *... nalectomy were not eligible. Patients who **had** prior radiotherapy for the treatment of*
4. *... tional. There were 12,798 patients who **did not have** radiotherapy and almost all ha*

O uso do verbo *have* chama atenção, pois no inglês geral, *have* seja usado na acepção de “tomar” ou “ingerir” (como em *have an icecream* ou *have some soup*). Nas três últimas linhas (2, 3 e 4), todas relacionando *have* e *radiotherapy*, o sentido passa a ser de recebimento de um tratamento, e o pronome relativo *who* faz com que a noção de passividade, de recebimento de algo feito por terceiros (no caso, o técnico de radiologia, que aplica a radiação) fique ainda mais evidente.

g. Emprego dos verbos *allocate*, *assign*, *attend*

Em algumas linhas de concordância observamos a utilização de uma estrutura diferente. Ao invés de dizer que os pacientes receberam determinado tratamento, os autores optaram por dizer que os pacientes foram designados para um grupo que recebeu tal tratamento, ou que eles frequentavam determinada clínica. Para expressar essa ideia, foram usados os verbos *allocate* (uma ocorrência), *assign* (duas ocorrências) e *attend* (quatro ocorrências, das quais apenas duas são apresentadas a seguir).

Exemplos:

1. *... follow-up, 97 (20%) study patients **allocated to** the conventional **therapy group** die*
2. *... an open-label randomised study. Patients **assigned to** the drug **therapy** assumed a*
3. *... free survival was observed. Patients initially **assigned** radiotherapy only could be su*
4. *... staging and CD4 counts for new patients **attending** antiretroviral **therapy clinic** Of*
5. *... s blood samples from new patients **attending** the adult antiretroviral **therapy clinic** fr*

Nas quatro ocorrências em que encontramos o uso do verbo *attend* no *corpus* ABROAD (duas delas ilustradas aqui), o objeto direto do verbo era algum tipo de clínica ou estabelecimento de saúde, para expressar a ideia de que o(s) paciente(s) frequenta o serviço de saúde e lá recebe o tratamento com regularidade.

h. Uso dos verbos *initiate, start, begin, use, continue, complete*

Em alguns casos, a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame) foi expressa utilizando-se verbos variados, como pode ser observado nas linhas de concordância abaixo:

Exemplos:

1. (*IRIS*) in HIV-infected patients **initiating** antiretroviral **therapy** (ART) results fro
2. end of March 2007, 95 674 patients **had started** free antiretroviral **therapy** in public
3. ght gain It is true that many patients who **begin** insulin **therapy** gain weight with impr
4. nonusers, whereas 9,001 patients **who continued** statin **therapy** had a highly signif
5. in RFI scores. Among the 5 patients who **were using** antidepressant **therapy**,
6. education, while 3 additional patients **completed** **therapy** after a 1-week delay followe
7. for stage II rectal cancer. One patient who **completed** protocol **therapy** was determin
8. trol rate is 32.4%. In the 83 patients **completing** induction chemoradio**therapy** and u

No total, foram observadas dez ocorrências dos verbos apresentados neste item, assim distribuídas: *initiate, start, begin* e *use*, uma ocorrência cada; *continue* e *complete*, quatro ocorrências cada. *Initiate, start, begin* expressaram a noção de que os pacientes estavam no início do tratamento, enquanto *continue* e *use* expressaram continuidade, e *complete* referiu-se à conclusão do tratamento.

i. Estruturas com *be considered for, eligible for*

Em ambos os casos, as frases passam uma noção de possibilidade de realização de determinado procedimento.

Exemplos:

1. *dissection. Especially in patients **considered for thrombolytic therapy** in acute str*
2. *in the present analysis. All patients **eligible for thrombolytic therapy** according to*

Ambas as estruturas foram utilizadas para expressar a noção de tratamento futuro, condição para ser submetido a um tratamento ou com planejamento para ser tratado com determinada terapia.

j. Estruturas com *to be given, to be subjected to*

Para ambos os casos, as frases são redigidas na voz passiva, deixando evidente a condição de receptor de determinado procedimento.

Exemplos:

1. *within the left atrium. The patient **was given** anticoagulant **therapy** and regained st*
2. *level of 128.3 µg/dl. The patient **was subjected to** chelation **therapy**, which was*

Ambos os tratamentos (terapia anticoagulante e terapia de quelação) são feitos com medicação controlada, que somente pode ser administrada para pacientes internados e/ou sob receita médica. Essa característica faz com que o uso de voz passiva seja mais indicado, pois o paciente de fato estará sendo o receptor do tratamento, e não apenas de prescrição desse medicamento.

k. Padrões com *Adjetivação*

Os autores que publicaram artigos nas revistas que compuseram o *corpus* ABROAD, além de utilizarem diversos verbos para expressar a noção de “submeter-se” a tratamento ou terapia, também utilizaram a adjetivação como estratégia linguística. O exemplo a seguir mostra uma ocorrência dessa estrutura.

Exemplo:

1. *colorectal cancer in chemotherapy-naive patients and patients pretreated with fluorouracil-based chemotherapy*

Nota-se, ainda, nessa mesma linha de concordância, uma segunda estrutura (sublinhada), que merece atenção. Observamos que, em uma mesma frase, há dois modos diferentes de expressar a noção de “submeter-se” a tratamento ou terapia, a adjetivação e o uso do verbo *treat*, com o prefixo *pre-*, e a preposição *with*.

3.5.1.1 Resumo dos padrões léxico-gramaticais encontrados para a coocorrência de *patient** e **therapy*

Com base nos achados resumidos nos Quadros 2 e 3, optamos por pesquisar a ocorrência simultânea de *patient** e **therapy*. Para exprimir a relação entre um tratamento médico a que foram submetidos pacientes para um estudo científico, foram empregados diferentes verbos, em vozes, modos e tempos verbais variados, conforme mostram os resultados acima. A Tabela 6 apresenta um resumo esquemático contendo, na primeira coluna, o verbo principal utilizado ou a estrutura gramatical empregada no textos do *corpus* ABROAD, com o número total de ocorrências entre parênteses; na segunda coluna os padrões léxico-gramaticais observados para a coocorrência, ou seja, a ocorrência simultânea em um horizonte predeterminado, das palavras *patient** e **therapy* (nos casos em que esses padrões foram observados); e, na terceira, o número de ocorrências de cada padrão ou estrutura correspondente ao significado atribuído a *submit* no *corpus* BRAZIL.

Grupo Verbal	PADRÃO LÉXICO-GRAMATICAL	n.
receive (93)	<i>patient(s) + received + (complementos) + *therapy</i>	37
	<i>patient(s) + receiving + (complementos) + *therapy</i>	25
	<i>patient(s) who received + (complementos) + *therapy</i>	14
	<i>patient(s) + had received + (complementos) + *therapy</i>	8
	<i>patient(s) + receive + (complementos) + *therapy</i>	3
	<i>patient(s) + was/were receiving/using + (complementos) + *therapy</i>	2
	<i>patient(s) who had received + (complementos) + *therapy</i>	2
	<i>patient(s) + has/have received + (complementos) + *therapy</i>	1
	<i>patient(s) + having received + (complementos) + *therapy</i>	1
treated with (47)	<i>patient(s) + treated with + (complementos) + *therapy</i>	37
	<i>patient(s) + was/were + treated with + (complementos) + *therapy</i>	9
	<i>patient(s) + being treated with + (complementos) + *therapy</i>	1
undergo (72)	<i>patient(s) + undergoing + (complementos) + *therapy</i>	44
	<i>patient(s) + underwent + (complementos) + *therapy</i>	20
	<i>patient(s) who underwent + (complementos) + *therapy</i>	4
	<i>patient(s) who had undergone + (complementos) + *therapy</i>	3
	<i>patient(s) who has/have undergone + (complementos) + *therapy</i>	1
Preposição (42)	<i>patient(s) + on + (complementos) + *therapy</i>	28
	<i>patient(s) + after + (complementos) + *therapy</i>	7
	<i>patient(s) was/were + on + (complementos) + *therapy</i>	3
	<i>patient(s) who is/are + on + (complementos) + *therapy</i>	3
	<i>patient(s) + being on + (complementos) + *therapy</i>	1
take	<i>patient(s) + TAKE/en/ing + (complementos) + *therapy</i>	4
have	<i>patient(s) + had + (complementos) + *therapy</i>	3
allocate (1); assign (2); attend (4)		7
initiate (2); start (2)		16
begin (1)		
use (3); continue (4)		
complete (4)		
be considered for (1)		2
elegible for. (1)		
to be given		2
to be subjected to		
Adjetivação		1

Tabela 6. Resumo esquemático dos resultados da busca de *patient** e **therapy* simultaneamente, no *corpus* ABROAD. A primeira coluna contém o verbo principal utilizado ou a estrutura gramatical empregada, com o número total de ocorrências entre parênteses; a segunda coluna apresenta os padrões observados e, a terceira, o número de ocorrências de cada padrão ou estrutura.

Pela análise dos dados citados anteriormente, resumidos na Tabela 6, podemos perceber que foram utilizados muitos verbos diferentes, ou estruturas gramaticais, para expressar a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame).

3.5.2. Padrões relacionados à coocorrência de *patient** e **ectomy*

Com base nos achados resumidos nos Quadros 2 e 3, optamos por pesquisar a ocorrência simultânea de *patient** e **ectomy*. Ao pesquisarmos a frequência de uso das palavras que possuíam o sufixo *-ectomy* (ou seja, dos termos contendo esse sufixo) entre as cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD, foram detectadas 218 ocorrências. Dessas, selecionamos manualmente as que correspondiam ao significado atribuído a *submit* no *corpus* BRAZIL, ou seja, à ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame).

a. Padrões contendo o verbo *undergo*

O verbo *undergo* foi o primeiro mais frequente ao pesquisarmos o uso de palavras contendo o sufixo **ectomy* no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas 33 ocorrências, distribuídas em seis padrões diferentes. Embora os padrões sejam formados por tempos verbais diferentes, em todos os casos analisados, as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s)* + *undergo* + (complementos) + **ectomy*

Exemplo:

*used as a temporizing measure before patients **undergo** decompressive craniectomy.*

Padrão: *patient(s)* + *undergoing* + (complementos) + **ectomy*

Exemplos:

esophageal cancer patients **undergoing** esophagectomy without prior neoadjuvant
 .The proportion of patients **undergoing** carotid endarterectomy who had optimal c
 % to 10% of BRCA+ patients **undergoing** prophylactic bilateral salpingo-oophorectomy

Padrão: patient(s) + underwent + (complementos) + *ectomy

Exemplos:

underwent coronectomy. Immediately postoperatively all patients were given
 The first patient **underwent** total laryngectomy for laryngeal cancer with
 One patient **underwent** a transhiatal esophagogastrectomy and 11 r

Padrão: patient(s) who + underwent + (complementos) + *ectomy

Exemplos:

also obtained from 16 patients **who underwent** thyroidectomy because of goitrous enl
 and steroids. One patient **who underwent** a bilobectomy developed pneumonia an
 xperimental group. Twenty-five patients **who underwent** total tonsillectomy (and were
 determined whether patients **who underwent** a carotid endarterectomy (n=252), cor
 tabase would not contain those patients **who underwent** modified radical mastectomy

Padrão: patient(s) + has/have undergone + (complementos) +
 *ectomy

Exemplos:

patients, 1 patient **has undergone** salvage prostatectomy after a positive p

Padrão: patient(s) + had undergone + (complementos) + *ectomy

Exemplos:

the records of 54 patients who **had undergone** prostatectomy, were referred for p
 5 patients **had undergone** modified radical mastectomy and 11 h

É interessante observar que os padrões acima foram utilizados com uma diversidade de procedimentos cirúrgicos (*bilobectomy, craniectomy, coronectomy, endarterectomy, esophagectomy, esophagogastrectomy, endarterectomy, laryngectomy, mastectomy* [duas ocorrências], *prostatectomy* [duas ocorrências], *salpingo-oophorectomy, thyroidectomy, tonsillectomy*). Essa diversidade revela que o uso do verbo *undergo* não fica restrito a uma área específica da medicina, sendo empregado em diferentes especialidades, desde a Ginecologia (*mastectomy; salpingo-oophorectomy*) e Urologia (*prostatectomy*), até especialidades como a Cardiologia/Cirurgia Vascular (*endarterectomy*) e Otorrinolaringologia (*tonsillectomy; laryngectomy*).

b. Padrões contendo as preposições *with, without* e a locução prepositiva *with and without*

As preposições *with, without* e a locução prepositiva *with and without* foram a segunda forma mais frequente ao pesquisarmos o uso de palavras contendo o sufixo **ectomy* no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas 13 ocorrências com *with*; oito com *without* e nove com a locução prepositiva *with and without*. Em todos os casos analisados, as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s) with/without + *ectomy*

1. patients **with** tonsillectomy and 7 (12.1%) without tonsillectom
2. patients **with** prior hypophysectomy or adrenalectomy were not eligible.
3. *nd without Tonsillectomy During the follow-up, 2 (3.70%)*
4. patients **with** tonsillectomy developed ESRF with mean renal s
5. patients **with** tonsillectomy and 70 **without** tonsillectomy in wh

Padrão: *patient(s) with and without + *ectomy*

Exemplos:

1. patients **with** and **without** tonsillectomy was 130 ± 50.3 month

É interessante observar que, na maioria dos casos (tanto de *with* quanto de *without* e da locução prepositiva *with and without*) os autores referiam-se ao termo *tonsillectomy*, o que nos leva a questionar se tais estruturas podem ser empregadas com outros tipos de intervenção cirúrgica, se este foi um achado isolado, ou se representa uma regra geral.

c. Padrões contendo o verbo *receive*

O verbo *receive* foi o terceiro mais frequente ao pesquisarmos o uso de palavras contendo o sufixo **ectomy* no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas oito ocorrências, distribuídas em três padrões diferentes. Embora os padrões tenham sido formados por tempos verbais diferentes, em todos os casos analisados, as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s) receiving + *ectomy*

Exemplo:

1. to solid food (B), patients **receiving** total tonsillectomy by electrocautery had

Padrão: *patient(s) received + *ectomy*

Exemplo:

1. ound that very few patients **received** carotid endarterectomy (less than 3% of the

Padrão: *patient(s) who received + *ectomy*

Exemplos:

1. *istics. Of the other 70 patients **who received** postorchidectomy radiation therapy*
2. *entiated thyroid carcinoma patients **who received** total thyroidectomy and high-dose*

d. Padrões contendo o verbo *treat*

O verbo *treat* foi observado ao pesquisarmos o uso de palavras contendo o sufixo **ectomy* no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Foram registradas cinco ocorrências. Em todas, os padrões foram formados pelo verbo na voz passiva ou na forma de particípio, exercendo a função sintática de complemento nominal, e sempre com a preposição *with*.

Padrão: *patient(s) treated with + *ectomy*

Exemplos:

1. *reated. A similar study involved patients **treated with** orchidectomy or combined HT*
2. *erentiated thyroid carcinoma patients **treated with** total thyroidectomy and high-dose*

Chama atenção o fato de *treated with* ter sido empregado para expressar a ideia de anterioridade de realização de determinado procedimento. Assim, essa estrutura assume um caráter de adjetivação, explicitando um subgrupo de pacientes, ou seja, aqueles que receberam certo procedimento. Assim como observamos em outros exemplos em que há adjetivação, a análise linguística das frases revela a presença de estruturas complexas, como em *differentiated thyroid carcinoma patients*.

e. Emprego do verbo *have*

O verbo *have*, como verbo principal, foi observado em algumas linhas de concordância, ao pesquisarmos o uso de palavras contendo o sufixo

*ectomy no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Em todos os casos observados, o verbo *have* estava no *Simple Past* e as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s) had + *ectomy*

Exemplos:

1. *rradiation. On B-24, all patients had lumpectomy and breast irradiation as above with*
2. *Erectile Function All patients had postprostatectomy, preradiation potency status e*

O uso do verbo *have* chama atenção, pois não seria um opção imediata para um autor brasileiro, muito embora, no inglês geral, *have* seja usado na acepção de “tomar” ou “ingerir” (como em *have an icecream* ou *have some soup*). Nos dois exemplos, o sentido é de recebimento de um procedimento cirúrgico, uma acepção diferente daquela empregada na língua geral.

f. Adjetivação

Em alguns casos, os procedimentos foram adjetivados (ou seja, formaram grupos nominais complexos do tipo substantivo com função de adjetivo + substantivo), para transmitir a ideia que os pacientes eram submetidos a eles. No total, foram três ocorrências dessa estrutura.

Padrão: **ectomy patient(s)*

Exemplos:

1. *Total tonsillectomy patients had longer mean time to no narcotic pain medicine*
2. *o lobectomy patients (6 days). There were no complications in patients*
3. *Postlumpectomy patients were treated with radiotherapy just to the r*

O uso de adjetivação é uma estratégia que parece tornar o texto mais compacto e claro, pois elimina o uso de preposições e verbos. Desse modo, o termo que representa o procedimento realizado passa a exercer a função de adjetivo (no caso do inglês, anteposto ao substantivo).

g. Padrões contendo a preposição *after*

O uso de preposições foi uma estratégia observada. Foram registradas duas ocorrências. Embora não sejam numericamente muitas, em ambos os casos as frases preservaram a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*. A seguir são apresentados dois exemplos dessas ocorrências.

Exemplos:

1. **After mastectomy**, hormone receptor-positive patients were given Tamoxifen
2. patients, mainly **after lumpectomy**. Postlumpectomy patients were treated with radiotherapy

É interessante notar que o exemplo 2 apresentado já havia sido citado em 3.3.1 (**Padrões contendo o verbo *treat***) e em 3.3.2 item f (**Adjetivação**). Assim, em duas frases seguidas do texto, o autor utilizou três escolhas diferentes (verbo *treat*, adjetivação e preposição - sublinhadas), para expressar a ideia de submeter-se a um procedimento. Vemos aí um exemplo da versatilidade dos autores dos textos do *corpus* ABROAD, que lançam mão de diferentes escolhas, para expressar um mesmo tipo de relação semântica.

h. Padrões contendo a estrutura *to be subjected to*

Nesse padrão, as frases foram redigidas na voz passiva, deixando evidente a condição de receptor de determinado procedimento.

Exemplo:

1. cGy were given. In patients **subjected to quadrantectomy**, the radiation consisted

3.5.2.1 Resumo dos padrões léxico-gramaticais encontrados ao pesquisar a coocorrência de *patient* e **ectomy*

Para realizar o sentido de pacientes submeterem-se a um procedimento médico para um estudo científico foram empregados diferentes verbos, em vozes, modos e tempos verbais variados, conforme mostrado nos resultados acima. Foi possível observar alguns padrões, em geral relacionados a determinados verbos, assim como já citamos para a pesquisa anterior (*patient(s) + *therapy*). A Tabela 7 apresenta, na primeira coluna, o verbo principal utilizado ou a estrutura gramatical empregada pelos autores dos textos do *corpus* ABROAD, com o número total de ocorrências entre parênteses; na segunda os padrões observados para a coocorrência de *patient** e **ectomy*; e, na terceira, o número de ocorrências de cada padrão correspondente ao sentido atribuído a *submit* no *corpus* BRAZIL.

Grupo Verbal	PADRÃO LÉXICO-GRAMATICAL	n.
undergo (33)	<i>patient(s) + undergoing + (complementos) + *ectomy</i>	9
	<i>patient(s) + underwent + (complementos) + *ectomy</i>	9
	<i>patient(s) who + underwent + (complementos) + *ectomy</i>	8
	<i>patient(s) + undergo + (complementos) + *ectomy</i>	2
	<i>patient(s) + has/have undergone + (complementos) + *ectomy</i>	1
	<i>patient(s) + had undergone + (complementos) + *ectomy</i>	1
	receive (13)	<i>patient(s) received + *ectomy</i>
<i>patient(s) who received + *ectomy</i>		5
<i>patient(s) receiving + *ectomy</i>		3
be treated with (5)	<i>patient(s) treated with + *ectomy</i>	5
have (2)	<i>patient(s) had + *ectomy</i>	2
subject to (1)	<i>patient(s) subjected to + *ectomy</i>	1
Grupo Nominal		
with (13); without (8)	<i>patient(s) with/without + *ectomy</i>	21
with and without (9)	<i>patient(s) with or without + *ectomy</i>	9
Adjetivação (3)	<i>*ectomy patient(s)</i>	3
Preposição (2)	<i>after + *ectomy</i>	2

Tabela 7. Resumo esquemático dos resultados da busca de *patient** e **ectomy* simultaneamente, no *corpus* ABROAD. A primeira coluna contém o verbo principal utilizado ou a estrutura gramatical empregada, com o número total de ocorrências entre parênteses; a segunda coluna apresenta os padrões observados e, a terceira, o número de ocorrências de cada padrão ou estrutura.

3.5.3 Padrões relacionados à coocorrência de *patient** e *procedure**

Com base nos achados resumidos nos Quadros 2 e 3, optamos por pesquisar a ocorrência simultânea de *patient** e *procedure**. Ao pesquisarmos a frequência de uso da palavra *procedure* (no singular e no plural) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD, foram detectadas 175 ocorrências. Dessas, selecionamos manualmente as que correspondiam ao significado atribuído a *submit* no *corpus* BRAZIL, ou seja, à ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame).

Das 175 ocorrências de *procedure**, entre as cinco palavras à direita ou à esquerda de *patient**, foi possível observar que o verbo mais comumente utilizado para expressar a noção de “submetido(s) a” foi *undergo* (conjugado em diferentes tempos, vozes e modos verbais). Entre as 175 ocorrências obtidas, 37 foram exemplos de uso de *undergo* das quais selecionamos, a seguir, as mais representativas.

a. Padrões contendo o verbo *undergo*

Padrão: *patient(s)* + *undergoing* + (complementos) + *procedure(s)*

Exemplos:

1. *in cirrhotic patients undergoing procedures that stress the cardiovascular system*
2. *perfect, patients undergoing these procedures benefit esthetically to a high degree*
3. *criteria, patients undergoing cardiac procedure remained significantly associated*
4. *teristics of patients undergoing different vascular procedures. We dropped from our*
5. *diatric oncology patients undergoing frequent invasive medical procedures and also*

Padrão: *patient(s) + underwent + (complementos) + procedure(s)*

Exemplos:

1. *period. If patients **underwent** the same **procedure** more than once during the period*
2. *later (day 7), patients **underwent** the same CT **procedure**, using the same time para*

Padrão: *patient(s) who + underwent + (complementos) + procedure(s)*

Exemplos:

1. *Patients who **underwent** the Norwood **procedure** had additional cavopulmonary s*
2. *ents who **underwent** more than one type of **procedure** during this 2-year period.*
3. *142 patients who **underwent** more than one type of vascular **procedure** during th*

Observamos que o verbo *undergo*, o mais recorrente ao pesquisarmos a coocorrência de *procedure* e *patients*, foi usado tanto para procedimentos clínicos (*CT procedure*) quanto cirúrgicos (*invasive medical procedures*), o que revela um emprego bastante diversificado, em textos de diferentes áreas da Medicina.

b. Padrões contendo o verbo *receive*

O verbo *receive* foi observado em duas ocorrências no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD. Em ambas, o padrão foi o mesmo, preservando a ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame), que no *corpus* BRAZIL foi expressa em diversos casos pelo verbo *submit*.

Padrão: *patient(s) + receiving + the procedure*

Exemplo:

1. *likelihood of privately insured patients **receiving** the **procedure** on an elective basis*

O uso de *receive* mostra claramente a situação de recebedor de um procedimento, mostrando que os pacientes foram submetidos ao procedimento, que foi realizado por um profissional que aplicou ou realizou o procedimento.

c. Frases contendo os verbos *do/perform*

Para ambos os casos, os verbos foram usados na voz passiva, e com o verbo auxiliar *to be* no *Simple Past*.

Exemplos:

1. *emergency **procedures** on patients already in hospital **were done** during course of*
2. *vascular **procedures** that **were performed** on 1680 patients. For the 105 patients*

Notamos que, no segundo exemplo, foi utilizada a preposição *on*, para mostrar que *patients* são os recebedores do procedimento, recuperando desse modo o significado de “submeter-se”. No primeiro exemplo, embora a frase esteja em ordem indireta, também observamos a preposição *on*.

3.5.3.1 Resumo dos padrões léxico-gramaticais encontrados ao pesquisar a coocorrência de *patient** e *procedure**

Para transmitir esse sentido, foram empregados diferentes verbos, em vozes, modos e tempos verbais variados. A Tabela 8 apresenta um resumo esquemático contendo, na primeira coluna, o verbo principal utilizado pelos autores dos textos do *corpus* ABROAD, com o número total de ocorrências entre parênteses; na segunda coluna os padrões observados para a coocorrência de *patient** e *procedure** (nos casos em que estes foram

observados); e, na terceira, o número de ocorrências de cada padrão ou estrutura com o significado atribuído a *submit* no *corpus* ABROAD.

Grupo Verbal	PADRÃO LÉXICO-GRAMATICAL	n.
<i>undergo</i> (37)	<i>patient(s) + undergoing + (complementos) + procedure(s)</i>	22
	<i>patient(s) + underwent + (complementos) + procedure(s)</i>	10
	<i>patient(s) who + underwent + (complementos) + procedure(s)</i>	5
<i>receive</i> (2)	<i>patient(s) + receiving + the procedure</i>	2
<i>do/perform</i> (2)	<i>procedure(s) + were done/performed on patient(s)</i>	2

Tabela 8. Resumo esquemático dos resultados da busca de *patient** e *procedure** simultaneamente, no *corpus* ABROAD. A primeira coluna contém o verbo principal utilizado ou a estrutura gramatical empregada, com o número total de ocorrências entre parênteses; a segunda coluna apresenta os padrões observados e, a terceira, o número de ocorrências de cada padrão ou estrutura.

3.5.4 Padrões relacionados à coocorrência de *patient** e *examination**

Com base nos achados resumidos nos Quadros 2 e 3, optamos por pesquisar a ocorrência simultânea de *patient** e *examination**. Ao pesquisarmos a frequência de uso da palavra *examination* (no singular e no plural) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD, foram detectadas 143 ocorrências. Dessas selecionamos manualmente aquelas que correspondiam ao significado atribuído a *submit* no *corpus* BRAZIL, ou seja, à ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame).

a. Padrão contendo o verbo *undergo*

Segundo a análise realizada, o único verbo utilizado para relacionar as palavras *patient(s)* e *examination(s)* com o sentido de “submeter-se” foi *undergo*, e em todos os casos registrados, tal verbo encontrava-se no

Simple Past. Foram encontradas 18 ocorrências na busca de *patient** e *examination** simultaneamente, no *corpus* ABROAD.

Padrão: *patient(s) + underwent + (complementos) + examination*

1. *patients (14.2%) underwent only undilated fundus examination, as the preserva*
2. *All patients underwent neurological examination at baseline as well as at three an*
3. *cted patients underwent B-scan examination, magnetic resonance imaging of the br*
4. *All patients underwent an MRI examination before randomization. MRI scanners o*
5. *All patients underwent a complete eye examination by a glaucoma specialist, with*

Uma observação bastante interessante foi que, ao analisarmos este padrão verbal, pudemos depreender um padrão nominal que foi utilizado de modo bastante consistente. Este último pode ser expresso da seguinte maneira:

Padrão: *a + (complementos) + examination.*

Exemplos:

1. *a detailed laboratory examination*
2. *a complete eye examination*
3. *a standard neurological examination*
4. *a detailed neurological examination*
5. *a complete ophthalmic examination*
6. *a full eye examination*
7. *a comprehensive ophthalmic examination*
8. *a complete ophthalmic examination*
9. *a complete physical examination*
10. *a full ocular examination*
11. *a conventional MRI examination*
12. *a full ophthalmic examination*
13. *a standard neurologic examination*

Como podemos notar, os elementos que compõem o complemento, no padrão supracitado, por sua vez, também seguiram um padrão bastante fixo, ou seja, o primeiro adjetivo expressando uma noção descritiva-

qualitativa do exame, utilizando termos como *detailed, complete, full, comprehensive, conventional, standard*; e o segundo fazendo uma referência direta ao campo da Medicina ou área do corpo em que o exame foi realizado, como em *neurological, ophthalmic, eye, ocular, physical, neurologic*. Desse modo, temos um padrão encaixado em outro.

3.5.5 Padrões relacionados à coocorrência de *patient** e **scopy*

Com base nos achados resumidos nos Quadros 2 e 3, optamos por pesquisar a ocorrência simultânea de *patient** e **scopy*. Ao pesquisarmos a frequência de uso das palavras que possuíam o sufixo *scopy* (ou seja, dos termos contendo esse sufixo) no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD, foram detectadas 28 ocorrências. Dessas selecionamos manualmente aquelas que correspondiam ao significado atribuído a *submit* no *corpus* BRAZIL, ou seja, à ideia de “submeter-se”, “passar por” ou “realizar” (tratamento, prova, teste ou exame).

Para transmitir esse sentido, foram empregados diferentes verbos, em vozes, modos e tempos verbais variados.

a. Padrões contendo o verbo *undergo*

Das 28 ocorrências de palavras contendo o sufixo **scopy*, entre as cinco palavras à direita ou à esquerda de *patient**, foi possível observar que o verbo mais comumente utilizado para expressar a noção de “submetido(s) a” foi *undergo* (conjugado em diferentes tempos, vozes e modos verbais). Entre as 28 ocorrências obtidas, oito foram exemplos de uso de *undergo* das quais selecionamos, a seguir, as mais representativas.

Padrão: *patient(s) + should undergo + (complementos) + *scopy*

Exemplo:

1. *patients with suspected achalasia **should undergo** upper gastrointestinal endoscopy*

Padrão: *patient(s) + underwent + (complementos) + *scopy*

Exemplo:

1. *all patients **underwent** an esophagogastroduodenoscopy with biopsy and esoph*
2. *The patient **underwent** rigid endoscopy under general anesthesia. After removal of*
3. *patients studied **underwent** surgical staging (mediastinoscopy and/or video as*

Padrão: *patient(s) + had undergone + (complementos) + *scopy*

Exemplo:

1. *of patients who **had undergone** bronchoscopy for diagnostic purposes after all ide*

Segundo pudemos observar, o verbo *undergo* foi utilizado com diferentes tipos de exame, tanto da área de Gastroenterologia (*esophagogastroduodenoscopy*), quanto de Cardiologia/Cirurgia Torácica (*mediastinoscopy*), evidenciando que, além de recorrente, o verbo apresenta uma incidência dispersa pelos diferentes textos e áreas da Medicina.

b. Padrões contendo o verbo *perform*

O verbo *perform* foi observado em algumas linhas de concordância, tanto na voz passiva (nos dois primeiros exemplos a seguir) quanto na voz ativa. Nessa última, teve como sujeito a expressão “*the patient's surgeon*”.

Exemplos:

1. *in one patient, specular microscopy **was performed** before and after the episode*
2. *and pelvis and repeat endoscopy **were performed on** the patients approximately*
3. *The patient's surgeon **performed** pretreatment endoscopy and was asked to doc*

Ao observarmos atentamente os exemplos de voz passiva, notamos que as preposições *in* (no primeiro caso) e *on* (no segundo) estabelecem a relação de *one patient* e *the patients*, respectivamente, como sujeitos da voz passiva de suas respectivas frases.

c. Padrões contendo os verbos *have*, *take*, *refer*

Observamos o uso dos verbos *have*, *take* (na forma de *to be taken*) e *refer* (na forma de *be referred for*) para se referirem a pacientes submetidos a determinado exame.

Exemplos:

1. *All patients **had** esophagogastroduodenoscopy (EGD) with biopsy of the primary l*
2. *-H or HSIL or if HPV DNA is positive, patients **should be referred for** colposcopy. If*
3. *patient **was taken to the operating room for** panendoscopy. Operative evaluatio*

Notamos que, no primeiro exemplo citado anteriormente o verbo *have* tem o sentido de “ser submetido a”. No segundo exemplo, o verbo *refer* poderia ser traduzido como “ser submetido” ou “ser encaminhado”. No terceiro, além da ideia de “ser submetido”, a locução verbal *was taken to* também pode ter o significado de “ser levado”, “ser transportado”.

3.5.5.1 Resumo dos padrões léxico-gramaticais encontrados ao pesquisar a coocorrência de *patient** e **scopy*

A Tabela 9 apresenta um resumo esquemático contendo, na primeira coluna, o verbo principal utilizado ou a locução verbal empregada pelos autores dos textos do *corpus* ABROAD, com o número total de ocorrências entre parênteses; na segunda coluna os padrões observados para a coocorrência de *patient** e **scopy* (nos casos em que estes foram observados); e, na terceira, o número de ocorrências de cada padrão ou

estrutura correspondente ao significado atribuído a *submit* no *corpus* BRAZIL.

Grupo Verbal	PADRÃO LÉXICO-GRAMATICAL	n.
<i>undergo</i> (8)	<i>patient(s) + underwent + (complementos) + *scopy</i>	6
	<i>patient(s) + should undergo + (complementos) + *scopy</i>	1
	<i>patient(s) + had undergone + (complementos) + *scopy</i>	1
<i>perform</i>	<i>*scopy + was/were + performed</i>	3
<i>have to be taken</i> <i>be referred for</i>		3

Tabela 9. Resumo esquemático dos resultados da busca de *patient** e **scopy* simultaneamente, no *corpus* ABROAD. A primeira coluna contém o verbo principal utilizado ou a estrutura gramatical empregada, com o número total de ocorrências entre parênteses; a segunda coluna apresenta os padrões observados e, a terceira, o número de ocorrências de cada padrão ou estrutura.

3.5.6 Padrões relacionados à coocorrência de *patient** e *test*

Com base nos achados resumidos nos Quadros 2 e 3, optamos por pesquisar a ocorrência simultânea de *patient** e *test*. Por fim, ao pesquisarmos a frequência de uso da palavra *test* no horizonte de cinco palavras à direita e à esquerda da palavra *patient(s)* no *corpus* ABROAD, foram observadas apenas duas ocorrências, e ambas utilizando o verbo *undergo* no tempo passado.

Padrão contendo o verbo *undergo*

Padrão: *patient(s) + underwent + (complementos) + test*

Exemplo:

1. *height and weight. If a patient **underwent** >1 metabolic stress **test**, only the initial st*
2. *screening 1. All patients **underwent** a visual acuity **test** under standardised lo*

Observamos novamente a ocorrência de *undergo* como o verbo utilizado para expressar o significado atribuído a *submit* pelos autores brasileiros.

3.6 Discussão sobre os dados obtidos

Serendipidade, acaso ou “sagacidade acidental”? Para a palavra inglesa *serendipity*, em português usam-se as formas *serendipity*, *serendipite* ou *serendipidade*. (SERRANO, 2007). Segundo o criador do termo, o escritor inglês Horace Walpole, a palavra deriva de Serendip, o antigo nome de Sri Lanka. Walpole escreveu em carta a um amigo: “*It was once when I read a silly fairy tale, called The Three Princes of Serendip: as their highnesses travelled, they were always making discoveries, by accidents and sagacity [...]*”. Um aspecto importante da definição original de Walpole é a “sagacidade”, a capacidade de fazer conexões entre fatos que aparentemente não teriam qualquer relação. Para Stoskopf (2005, p. 332-7), o desejo de empregar seus “esforços neuronais” em questões que possam não ter uma relevância imediata é o traço que distingue um “pesquisador tecnicamente proficiente” e um “cientista criativo e estimulante” (tradução nossa). Para Stoskopf, seria a manutenção vitalícia dessa “curiosidade básica” que caracterizaria os cientistas realmente grandes, evitando a “indolência intelectual” que tão comumente, segundo ele, acompanha o reconhecimento profissional.

Outros autores também reconheceram a importância da serendipidade nas descobertas científicas. Campbell (2005, p. 353) considera que, embora a aplicabilidade do termo serendipidade às descobertas farmacológicas possa ser incerta, o papel desempenhado pelas descobertas “ao acaso” precisaria ser reconhecido e considerado bem-vindo. Davisson (2005, p.338), na mesma edição da revista publicada pelo *Institute of Laboratory Animal Resources*, afirmou que o profissional “alerta” e “astuto” será capaz de descobrir, pela observação cuidadosa durante as operações de rotina, aspectos importantes e conseguirá explorá-los de modo a criar oportunidades de novas descobertas, de algo que não estava sendo

originalmente buscado. Morton & Swindler (2005) expandem ainda mais o conceito, considerando a serendipidade como um “aspecto fundamental” das pesquisas básicas, e um desfecho valioso e corriqueiro em casos de “observação informada”. Ou seja, Morton & Swindler consideram que as pesquisas não podem prescindir da serendipidade, e assim conferem a ela um papel relevante e indispensável à atividade de pesquisador.

Mesmo nas atividades clínicas diárias, a serendipidade vem sendo cada vez mais reconhecida por pesquisadores do mundo todo como uma importante ferramenta diagnóstica (ALZHRANI *et col.*, 2008; KULAMBIL PADINJAKARA, ASHAWESH & PATEL, 2008), terapêutica (CAGIANO *et col.*, 2008; DASGUPTA, 2009) e de definição de rumos profissionais (SUSSMAN, 2008). Entretanto, Saalfrank & Scheurer Maid (2008) enfatizam que a serendipidade pode dar início a diversas descobertas, mas sua materialização deve ser baseada em um planejamento racional e embasada em conhecimentos científicos.

Nas ciências humanas, por sua vez, também há estudos que realçam a importância da serendipidade. No caso de grandes conjuntos de textos (como no caso dos nossos *corpora*). Stevenson (2008) procura explorar a importância das referências cruzadas e das conexões explícitas (citações, referências) e implícitas entre livros e artigos científicos. Segundo o autor, os textos em formato eletrônico podem fornecer informações de um modo muito mais rápido e eficiente do que o mesmo texto, em formato de papel. Além disso, há uma maior possibilidade de localizar estruturas textuais, desde que se utilizem ferramentas específicas para tanto (no nosso caso, as ferramentas do programa *WordSmith* foram concebidas especificamente para esse tipo de pesquisa) e que se lance mão de uma “serendipidade estruturada”. A estruturação, segundo o autor, refere-se à maneira ordenada e predeterminada de se buscar informações em um texto ou conjunto de textos. Assim, mais uma vez, notamos que os autores, ao falar de serendipidade, sempre enfatizam a importância do planejamento e estudo prévios, abrindo caminhos para a busca, e de um raciocínio racional e científico embasando a pesquisa.

Uma autora que se dedica a pesquisar a influência da serendipidade ligada aos estudos com *corpus* é Bernardini (2000). Essa autora italiana

considera sempre as limitações das ferramentas computacionais na pesquisa em *corpus*, e reconhece que as descobertas feitas com base em *corpora* muitas vezes são limitadas pelas próprias ferramentas utilizadas para sua exploração. Segundo Bernardini, trabalhar com *corpus* pode ser uma “fonte rica” de informações linguísticas, e quem pesquisa em *corpora*, seja aluno ou profissional, irá se beneficiar ainda mais se realizar atividades que estejam reforçadas pela serendipidade, desde que tenham um embasamento científico e teórico.

Assim, com base na literatura apresentada acima, percebemos que a serendipidade pode e deve ser levada em consideração nas pesquisas de *corpus*. Afinal, pesquisadores das mais variadas áreas, e de diferentes partes do mundo têm atribuído valor a esta estratégia de pesquisa. Entretanto, em todos os textos citados, fica evidente a importância de se ponderar os achados, refletir sobre eles e analisar cada nova perspectiva à luz dos conhecimentos adquiridos anteriormente, das informações fundamentadas em bases científicas e em descobertas guiadas pelo estudo aprofundado dos aspectos estruturais da matéria que se pretende analisar.

Do ponto de vista da Linguística de *Corpus*, o valor da serendipidade fica evidenciado ao estudarmos as listas de palavras, linhas de concordância, n-gramas e colocações e fazermos “descobertas” relacionadas ao funcionamento de determinada língua, variante, autor ou texto, pois, embora lidando com dados produzidos pela manipulação de *corpus* por ferramentas eletrônicas, o pesquisador sempre fundamenta suas reflexões em um estudo prévio, impulsionando seus achados na constatação de coocorrências e/ou recorrências que muitas vezes são apontadas pelas ferramentas computacionais, mas que aparentam ocorrer “casualmente”. Wilkinson (2007, p.1) afirma que a pesquisa em *corpus* permite que se descubram informações imprevisíveis, de modo que o usuário poderá observar usos desconhecidos ou inovadores, em uma linha de concordância, e terá condições de explorar novas perspectivas e seguir trilhas a partir de descobertas aparentemente ocasionais.

Na presente pesquisa, tendo em vista os resultados apresentados, podemos observar que, para um mesmo verbo utilizado nos textos do *corpus*

BRAZIL (*submit*), foram empregados diversos verbos, preposições e substantivos no *corpus* ABROAD. Em alguns casos, os autores de língua inglesa que publicam artigos em revistas médicas do exterior optaram por usar preposições (74 ocorrências) para expressar a ideia de “submeter-se”, entre elas podemos salientar *after*, *on*, *with*, *without* e “*with and without*”. As escolhas que lançaram mão de adjetivação e preposições são resumidas na Tabela 10.

Estratégia	PADRÃO LÉXICO-GRAMATICAL	n.
Preposição (74)	<i>patient(s) + on + (complementos) + *therapy</i>	28
	<i>patient(s) with/without + *ectomy</i>	21
	<i>patient(s) + after + (complementos) + *therapy/*ectomy</i>	9
	<i>patient(s) with or without + *ectomy</i>	9
	<i>patient(s) was/were + on + (complementos) + *therapy</i>	3
	<i>patient(s) who is/are+ on + (complementos) + *therapy</i>	3
	<i>patient(s) + being on + (complementos) + *therapy</i>	1
Adjetivação (4)	<i>+ *therapy/*ectomy patient(s)</i>	4

Tabela 10. Resumo esquemático das escolhas empregadas pelos autores de textos do *corpus* ABROAD para expressar o sentido de *submit* através do uso de sintagmas nominais. A primeira coluna contém a estrutura gramatical empregada, com o número total de ocorrências entre parênteses; a segunda coluna apresenta os padrões observados e, a terceira, o número de ocorrências de cada padrão ou estrutura.

As escolhas relacionadas ao uso de diferentes verbos são apresentadas na Tabela 11 (pág. 99). Nela, a primeira coluna contém o verbo principal utilizado ou a estrutura gramatical empregada, com o número total de ocorrências entre parênteses; a segunda coluna apresenta os padrões observados e, a terceira, o número de ocorrências de cada padrão ou estrutura. Entre as escolhas empregadas, o verbo *undergo* (em diferentes tempos e vozes) foi o preponderante, com 170 ocorrências, seguido por *receive* (108 ocorrências). Associado a uma preposição (*with*), o verbo *treat* na voz passiva também foi uma escolha frequente no *corpus* de referência (52 ocorrências).

undergo (170)	patient(s) + undergoing + complementos + *therapy/*ectomy/procedure(s)	75
	patient(s) + underwent/ who has/have undergone + complementos + *therapy/*ectomy/ procedure(s)/examination*/*scopy/test	67
	patient(s) who underwent + complementos + *therapy/*ectomy/ procedure(s)	17
	patient(s) who had undergone + complementos + *therapy/*ectomy/*scopy	5
	patient(s) + undergo + complementos + *ectomy	2
	patient(s) + should undergo + complementos + *scopy	1
receive (108)	patient(s) + receive/ having received + complementos + *therapy	4
	patient(s) + receiving + complementos + *therapy/ *ectomy/ the procedure	30
	patient(s) + was/were receiving/using + complementos + *therapy	2
	patient(s) + received/ who received + complementos + *therapy/*ectomy	61
	patient(s) + has/have received/ had received/ who had received + complementos + *therapy	11
treated with (52)	patient(s) + being treated with/ was/were + treated with + complementos + *therapy	10
	patient(s) + treated with + complementos + *therapy/*ectomy	42
take (4)	patient(s) + take + complementos + *therapy	4
have (4)	patient(s) + had + complementos + *therapy/*ectomy/*scopy	6
initiate (2), start (2), begin (1), use (3), continue (4), complete (4)	+ *therapy ou + procedure ou + nome da terapia ou medicamento	16
allocate (1), assign (2), attend (4)	+ *therapy + clinic	7
to be taken; be referred for	*scopy	2
perform (4)	*scopy/procedure(s) + was/were + performed	4
do (1)	procedure(s) + were done/performed on patient(s)	1
be considered for (1); eligible for (1):	+ *therapy	2
to be given (1); to be subjected to (1)	+ *therapy	2
subject to (1)	patient(s) subjected to + *ectomy	1

Tabela 15. Resumo esquemático das escolhas do corpus ABROAD para expressar o sentido de *submit* através do uso de diferentes sintagmas verbais.

Outro aspecto que chama a atenção foi a variedade de estruturas linguísticas utilizadas, com o uso de 24 verbos diferentes (*undergo, receive, treat, take, have, initiate, start, begin, use, continue, complete, allocate, assign, attend, to be taken, be referred for, perform, do, be considered for, eligible for, to be given, to be subjected to*) preposições (já citadas), adjetivação (quatro ocorrências), totalizando 47 estruturas e padrões diferentes. Esta variedade de opções pode ser melhor observada no Gráfico 1 (pág. 101).

Os autores dos textos do *corpus* ABROAD lançaram mão de um número grande de padrões léxico-gramaticais para expressar o sentido do item lexical estudado, chegando a utilizar até três estruturas diferentes em uma mesma frase.

Percebemos que o conhecimento da fraseologia típica de um registro (o médico-científico) e de um gênero (o artigo científico) pelos autores parece ser uma característica que se distribui pelo texto. Dito de outro modo, os autores que utilizam padrões léxico-gramaticais mais adequados ao gênero e tipo textual parecem usá-los repetidas vezes em um mesmo período ou parágrafo. O que revela que esse autores têm um domínio mais elevado do idioma no qual estão se expressando.

Em relação à metodologia empregada, à luz dos resultados obtidos, observamos que as pesquisas *corpus driven* são valiosas, pois tal metodologia serviu para encontrar padrões até então não conhecidos, tendo em vista o ponto de partida, que foi o sentido de “submeter-se”. Além disso, não havia na literatura indicação de que o uso excessivo de *submit* seria um aspecto recorrente na escrita médico-científica brasileira, e nosso estudo acabou revelando esse achado.

Os achados sustentam os pressupostos da Linguística de *Corpus*, pois os estudiosos da LC (SINCLAIR, 1991; BERBER SARDINHA, 2000a; entre outros) consideram que o significado encontra-se nos padrões. Os seus padrões, portanto, não são meras combinações de palavras. Eles carregam sentidos distintos. O estudo desses padrões pode revelar informações sobre o nível de domínio de um idioma pelo autor do texto.

Percebemos que não houve uma relação um para um entre os padrões escritos por grupos diferentes de autores. As revistas científicas brasileiras usam determinados padrões e as estrangeiras outros. De certo modo, isso está previsto indiretamente na literatura, pois em nossa experiência com tradução e nos estudos de tradução com *corpus* notamos que não há correspondência um para um entre texto traduzido e texto original. Embora os textos do *corpus* de estudo não sejam necessariamente traduzidos, podem ter sido, pois sabemos por nossa experiência na área que alguns autores primeiramente redigem seus artigos em português, para depois traduzi-los para o inglês.

A partir deste estudo *corpus driven* foi possível constatar que há diferença entre os brasileiros e os estrangeiros, quando escrevem o mesmo gênero e o mesmo registro. As escolhas são diferentes. Isso significa, segundo Halliday (BERBER SARDINHA, 2004) quando isso ocorre, houve '*resetting of the system*'. Esse *resetting* (reajuste) mostra que os brasileiros e os estrangeiros estão operando em sistemas diferentes. O sistema linguístico dos brasileiros é um (com suas probabilidades) e o dos estrangeiros é outro. Sendo assim, acreditamos que seja importante determinar quais seriam os pontos em que o reajuste do sistema levou o texto a apresentar escolhas de baixa frequência nos textos considerados como padrões de prestígio. Afinal, a proximidade idiomática com esses padrões de prestígio faz com que os textos produzidos possam se equiparar em termos linguísticos aos textos publicados em revistas médicas do exterior. Aparentemente estamos observando a mesma língua (inglês), em dois *corpora* distintos, mas na verdade as diferentes escolhas e opções por padrões léxico-gramaticais diferentes fazem com que os textos de um e de outro *corpus* comportam-se como "línguas diferentes", cada uma com seu código.

Qual seria a origem desses vários padrões? De acordo com Hoey (2005), os padrões residem na nossa mente e são incorporados a nosso patrimônio intelectual a partir da exposição constante à língua (seja escrita ou oral). A mente recorta e guarda esses padrões na mente. No caso das revistas estrangeiras, isso ficou evidente: há vários padrões utilizados nos

textos. No caso das revistas brasileiras, a escassez de padrões para um mesmo sentido indica que os autores não possuem *primings* variados para esse sentido, o que nos leva a concluir que não foram expostos ou não foram capazes de perceber os padrões e sistematizá-los na mente.

Supondo-se que os autores que escrevem esses artigos leiam bastante em inglês (que, segundo nossa experiência na área, parece ser o que ocorre em muitos casos), então o achado de poucos padrões não seria uma questão de falta de exposição ao inglês escrito – já que muitos médicos assinam revistas científicas da área, ou têm acesso a elas em seus locais de trabalho, e costumam ler artigos em inglês com uma frequência que, embora ainda não tenha sido determinada estatisticamente em nenhum estudo científico, parece estar acima da média observada para outros grupos profissionais. A falta ou escassez de uso dos padrões léxico-gramaticais pode, então, ser resultado da falta de consciência da existência e da importância desses padrões.

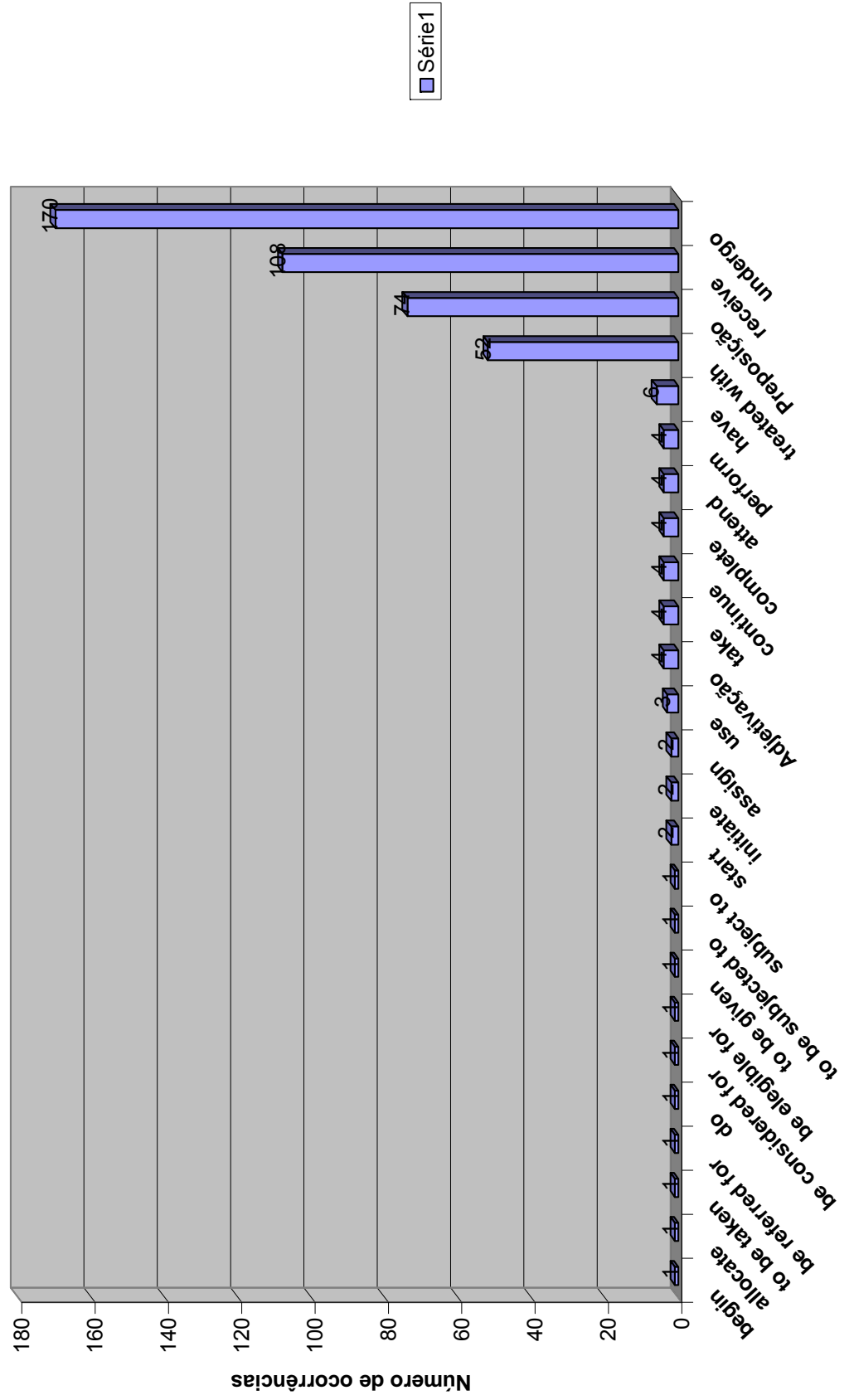
É preciso mudar essa situação, pois escrever com os padrões semelhantes aos empregados nas revistas estrangeiras significa escrever com o mesmo sistema (Halliday, apud BERBER SARDINHA, 2006). Ao escrever com o mesmo sistema, o texto passa a ser melhor apreciado pelo outro grupo (no caso, o dos médicos estrangeiros); com isso, aumentam as chances de serem publicados nos periódicos internacionais de prestígio.

E como mudar essa situação? É preciso ser ensinado ou ser exposto a esses padrões, para que o indivíduo fique '*primed*' a eles. (BERBER SARDINHA, 2009, comunicação pessoal). No caso de estrangeiros, esses *primings* não surgem espontaneamente, com a mera exposição contínua, pois se o indivíduo não tiver consciência deles, se não souber perceber e dar valor a eles, dificilmente irá incorporar os *primings* e, possivelmente, não irá usá-los. Principalmente se o médico brasileiro tiver uma formação de inglês baseada no conhecimento e domínio principalmente da gramática, ele dificilmente irá perceber os padrões e não ficará *primed*.

Observo, em minha experiência profissional ministrando cursos de redação de artigos científicos em inglês para médicos e profissionais da área da saúde, que esses profissionais querem melhorar sua maneira de redigir, estão motivados a fazê-los, mas dificilmente encontram um profissional

disposto a lhes ensinar algo além da gramática, dos tempos verbais e dos exercícios para memorização e tradução de vocabulário específico. Percebo que muitos profissionais chegam desanimados, acreditando que não há uma solução para seus problemas de escrita, já que diversos deles já têm proficiência em língua inglesa (exigida pela maioria das faculdades de Medicina, para acesso à carreira universitária). Com um curso desenvolvido de modo a instrumentalizar os alunos no uso de algumas ferramentas computacionais, noções básicas de Linguística de *Corpus* e exemplos práticos de como as linhas de concordância podem revelar muito mais do que um livro de gramática um dicionário poderia oferecer, é bastante gratificante observar que a metodologia *corpus driven*, além de ser útil para pesquisadores, também pode ajudar a solucionar problemas de autores que queiram dar a seu texto maior proximidade idiomática com os textos considerados como padrões de prestígio. Pelo uso consciente de alguns recursos da LC, os médicos que pretendem redigir artigos em inglês, para publicarem em revistas científicas nacionais e internacionais podem adquirir conhecimento de padrões léxico-gramaticais de uso frequente nos textos publicados em inglês nos EUA e Europa, tornando sua escrita mais próxima desse padrão de excelência.

Gráfico 1. Estratégias usadas no corpus ABROAD para expressar a idéia de "submeter-se"



■ Série1

CONSIDERAÇÕES FINAIS

O objetivo geral da pesquisa apresentada nesta dissertação foi identificar os padrões léxico-gramaticais característicos de textos médicos escritos em inglês publicados em revistas do Brasil e, a seguir, procurar determinar se tais padrões são recorrentes em textos de mesmo tipo textual (no caso, artigos científicos publicados em revistas científicas) e público alvo (profissionais de saúde) similar, publicados nos Estados Unidos e Europa. Usamos dois *corpora* comparáveis (BRAZIL, *corpus* de estudo e ABROAD, *corpus* de referência), com aproximadamente cinco milhões de *tokens* cada.

O presente estudo também analisou aspectos característicos dos textos médicos publicados em inglês no Brasil, levantando quais seriam os padrões correspondentes, nos textos publicados em revistas médicas do exterior, usados para realizar sentidos semelhantes. A partir dessa análise, observamos que havia 2.504 trigramas exclusivos dos textos publicados no Brasil, apontando para um problema de interferência da língua materna (português) que afetaria o momento da produção das frases em inglês.

O objetivo específico deste estudo foi determinar como os médicos brasileiros usaram o verbo *submit*, uma tradução *prima facie* da palavra “*submeter*”, e detectar quais estruturas gramaticais foram usadas no *corpus* ABROAD para expressar o mesmo significado.

A análise dos dados obtidos trouxe os seguintes achados: (i) o verbo *submit* foi usado nas revistas publicadas nos EUA e na Europa com o sentido de “enviar (um texto) para análise” ou “enviar para publicação”, enquanto no *corpus* BRAZIL o referido verbo foi utilizado com o sentido de “submeter-se” ou “encaminhar” um paciente para fazer um exame, teste, procedimento ou tratamento. Além disso, foram observadas apenas quatro ocorrências de *submit* no *corpus* ABROAD, em comparação com 1.111 ocorrências do verbo no *corpus* BRAZIL, revelando um uso excessivo. (ii) As escolhas encontradas no *corpus* ABROAD para expressar o sentido de “submeter-se a um exame, teste ou terapia” foram: uso de verbos (tais como

undergo; receive; take; have; initiate, start, begin, use, continue, allocate, assign, attend, complete; to be taken, be referred for; perform, do, be considered for, eligible for, to be given, to be subjected to, subject to) em diferentes tempos, vozes e modos; uso de adjetivação; uso de preposições (*after, on, with, without*). (iii) De todas as escolhas feitas para expressar a ideia de “submeter-se a um exame, teste ou terapia,” a utilizada com maior frequência pelos autores dos textos do *corpus* ABROAD foi o verbo *undergo*, em diferentes tempos verbais, vozes e modos. (iv) O uso de grupos nominais (adjetivação, por exemplo) frequente no *corpus* de referência. A escolha por um grupo nominal torna mais complexa a compreensão do sentido, pois há um grande número de relações existentes entre os substantivos de um grupo nominal (causa, fonte, material, etc). (BIBER, 1999) Sendo assim, a presença de escolhas formadas por grupos nominais parece revelar um domínio maior do idioma.

De posse das análises dos dados selecionados, podemos concluir que: (i) os autores brasileiros fizeram um uso excessivo e inadequado do verbo *submit*. (ii) Para um mesmo verbo (*submit*) utilizado nos textos do *corpus* BRAZIL, foram observados diversos verbos e estruturas correspondentes no *corpus* ABROAD, revelando que os autores que publicaram no exterior lançam mão de uma maior gama de escolhas linguísticas para expressar uma mesma ideia. (iii) Caso um indivíduo que esteja escrevendo um texto em inglês precise traduzir o verbo “submeter”, a tradução mais segura deveria ser *undergo*. (iv) O uso de adjetivação é uma estratégia que parece tornar o texto mais compacto, pois elimina o uso de preposições e verbos. Desse modo, o termo que representa o procedimento realizado passa a exercer a função de adjetivo (no caso do inglês, anteposto ao substantivo) com o uso de um número menor de palavras, e sem precisar lançar mão de verbos.

Como qualquer pesquisa, a nossa também possui limitações, sendo a maior delas o tempo exíguo para a conclusão de uma análise de um conjunto tão numeroso de informações. Razão pela qual, ao concluirmos esta dissertação, temos a sensação de que seria o melhor momento para

começar outra, e mais outra, enfocando diversos aspectos que surgiram e outros tantos que surgirão a cada nova incursão aos dados.

Aspectos como o uso de voz passiva nos textos científicos, o emprego do verbo *have* em diferentes acepções, as especificidades de sentido dos diferentes padrões, ao compararmos língua geral e língua de especialidade, e o uso de adjetivação são aspectos que se mostraram relevantes, e que merecem estudos futuros mais detalhados.

O uso de diferentes escolhas (*treated, undergo*, adjetivação, entre outras) certamente pode refletir diferenças de sentido nos textos. Presumimos que existam especificidades de sentido que precisam ser analisadas em estudos posteriores.

Quanto às perspectivas futuras, acreditamos que o uso da voz passiva nos textos científicos em inglês precise ser analisado em maior profundidade, assim como a relação entre o uso de adjetivação e o nível de proficiência linguística do autor de textos médicos.

Com a pesquisa apresentada nesta dissertação, esperamos ter dado uma pequena contribuição para os estudos de Linguística de *Corpus* e para os Estudos de Tradução. Esperamos também ter contribuído para um aumento na compreensão das estruturas utilizadas pelos profissionais que publicam em inglês em revistas médicas dos Estados Unidos e da Europa. Como os artigos analisados eram das áreas de Medicina Humana e Odontologia, esperamos que nossas conclusões sejam valiosas para articulistas, da área da saúde, tradutores, linguistas, terminólogos e jornalistas que trabalhem com redação de textos médicos e de áreas afins.

Referências

- ALAMBERT, E. G. S. **Panorama histórico da linguística aplicada**. São Paulo, inédito, 2006.
- ALAMBERT, E. G. S. **Uma tradução premiada sob a perspectiva da Linguística de Corpus**. São Paulo, SP, 2008. Originalmente apresentada como dissertação de mestrado, Pontifícia Universidade Católica de São Paulo, 2004. Disponível em: <http://www.pucsp.br/pos/lael/lael-inf/def_teses.html>. Acesso em: 24 jan. 2009.
- ALZHRANI, A.S. AL-KHALDI, N.; SHI, Y.; AL-RIJJAL, R. A.; ZOU, M.; BAITEI, E. Y.; AMIN, T. Diagnosis by serendipity: Cushing syndrome attributable to cortisol-producing adrenal adenoma as the initial manifestation of multiple endocrine neoplasia type 1 due to a rare splicing site MEN1 gene mutation. **Endocrine Practice**, EUA, v. 14, n. 5, p. 595-602, Jul-Aug 2008.
- AMERICAN MEDICAL ASSOCIATION. Iverson C. **Manual of style: a guide for authors and editors**. Philadelphia: Williams & Wilkins; 1998.
- ANGELL, M. Publish or perish: a proposal. **Annals Internal Medicine**. EUA, v.104, n. 2, p. 261-2, fev.1986.
- ANHANGUERA EDUCACIONAL. **Grade Curricular Pós Graduação Tradução Inglês Português**. São Paulo, 2008. Disponível em <http://ww4.unianhanguera.edu.br/posgraduacao/hotsite/descritivo_curso_ingles_portugues.php>. Acesso em 20 fev. 2009.
- BAKER, M. Corpora in Translation Studies: an overview and some suggestions for future. **Target**, v. 7, n. 2, p. 223-43, 1995.
- BAMBERG, B. What makes a text coherent? **College Composition and Communication**, v. 34, n. 4, p. 417–29, 1983.
- BENNETT D.M., TAYLOR D.M. Unethical practices in authorship of scientific papers. **Emergency Medicine**, Melbourne, Austrália, v.15, p. 263-70, 2003.
- BERBER SARDINHA, A. P. Linguística de Corpus: Histórico e Problemática. **Delta**, São Paulo, v.16, n. 2, p. 323-67, 2000a.
- BERBER SARDINHA, A. P. O que é um corpus representativo? **DIRECT Papers**, São Paulo, 2000b. Disponível em <<http://www.direct.f2s.com>>, Acesso em 23.jan.2008.

BERBER SARDINHA, T. Usando *WordSmith Tools* na investigação da linguagem. **DIRECT Papers**, São Paulo, v. 40, 1999. Disponível em <<http://74.125.47.132/search?q=cache:w4Me1s5HrYJ:www2.lael.pucsp.br/direct/DirectPapers40.pdf+sardinha+type+token+n%C3%BAmero&cd=5&hl=pt-BR&ct=clnk&gl=br>> Acesso em 20 set. 2008.

BERBER SARDINHA, T. **Linguística de corpus**. Barueri: Editora Manole, 2004.

BERBER SARDINHA, T. Halliday's probability resetting and metaphor use: Comparing a specific genre to general language. In: 33rd International Systemic Functional Congress, 2007, São Paulo. **Proceedings of the 33rd International Systemic Functional Congress**, São Paulo: PUCSP, v. 1. p. 977-988, 2006.

BERNARDINI, S. Exploring new directions for discovery learning. In KETTEMANN B. & MARKO G. (Org.) **Teaching and learning by doing corpus linguistics**. Papers from the Fourth International Conference on Teaching and Language Corpora, 19-24 jul. 2000. Amsterdam & Atlanta, Georgia: Rodopi, 2002. p. 165-182.

BERNARDINI, S. Serendipity expanded: Exploring new directions for discovery learning. In **Fourth International Conference on Teaching and Language Corpora**, Graz. Disponível em <<http://www-gewi.kfunigraz.ac.at/talc2000/Dokumente/abstracts/>> Acesso 18 jan.2009.

BERNARDINI, S. Systematising serendipity: Proposals for concordancing large corpora with language learners. In BURNARD, L. & McENERY T. (org.) **Rethinking language pedagogy from a corpus perspective**. Frankfurt am Main: Peter Lang, 2000. p. 225-234.

BIBER, D.; CONRAD S.; CORTES V. If you look at ... lexical bundles in academic lectures and textbooks. **Applied Linguistics**, Oxford, v. 25, n. 3, p. 371-405, 2004.

BIBER, D.; CONRAD, S. Lexical bundles in conversation and academic prose. In: HASSELGARD, H.; OKSEFJELL, S. (eds.). **Out of corpora: Studies in honor of Stig Johansson**, Amsterdam: Rodopi, 1999. p. 181-9.

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, 2003. p. 71-92.

BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. **Grammar of Spoken and Written English** Longman , Harlow: Pearson Education Ltd., 1999.

BOWKER, L. Using specialized monolingual native language corpora as a translation resource: a pilot study. **Meta**, v. 43, n. 4, p. 631-51, 1998.

BRAILE, D.M.; BRANDAU, R.; MONTEIRO, R. A importância da indexação para as revistas científicas: [editorial]. **Revista Brasileira de Cardiologia Intensiva**. Brasil, v.15, n. 4, p. 341-2, out.-dez..2007.

CAGIANO R. BERA, I.; VERMESAN, D.; FLACE, P.; SABATINI, R.; BOTTALICO, L.; AUTERI, P.; SANTACROCE, L. Psoriasis disappearance after the first phase of an oncologic treatment: a serendipity case report. **La Clinica Terapeutica**, Itália, v. 159, n. 6, p. 421-5, Nov-Dez 2008.

CAMPBELL, W.C. Serendipity and new drugs for infectious disease. **ILAR Journal**, Washington, EUA, v. 46, n. 4, p. 352-6, 2005.

CANDIDO Jr., A. **Criação de um Ambiente para o Processamento de Córpus de Português Histórico**. São Carlos, fev. 2007. Originalmente apresentada para o Exame de Qualificação para obtenção do título de Mestre em Ciências de Computação e Matemática Computacional pelo Instituto de

CASTRO R. C. F., FERREIRA M. C. G., VIDILI A. L. Periódicos latinoamericanos: avaliação das características formais e sua relação com a qualidade científica. **Ciência da Informação**, Brasília, DF; v. 25, n. 3, p. 357-67, set/dez. 1996.

CELANI, M.A.A. Afinal, o que é Lingüística Aplicada? In: PASCHOAL, M. S. Z. de & CELANI, M.A.A.(Orgs.) **Lingüística Aplicada: da Aplicação da Lingüística à Lingüística Transdisciplinar**. São Paulo: Educ. p. 15-23, 1992.

Ciências Matemáticas e de Computação -ICMC-USP. Disponível em:
<<http://www.lettras.etc.br/ebral/quali-arnaldo.pdf>> Acesso 15 jan. 2009.

COELHO, P.M.Z.; ANTUNES, C.M.F.; COSTA, H.M.A.; KROON, E.G.; SOUSA LIMA, M.C.; LINARDI, P.M. The use and misuse of the "impact factor" as a parameter for evaluation of scientific publication quality: a proposal to rationalize its application. **Brazilian Journal of Medical and Biological Research**, v. 36, n. 12, p.1605-12, dez. 2003.

CONRAD, Susan M. Investigating Academic Texts with Corpus-based Techniques: an Example from Biology. **Linguistics and Education**, Amsterdam, v. 8, n. 3, p. 299 – 326, set. 1996.

CORTES, V. Lexical bundles in published and student disciplinary writing: Examples from history and biology. **English for Specific Purposes**, v. 23, n. 4, p. 397–423, 2004.

CORTES, V. Teaching lexical bundles in the disciplines. A example form a writing intensive history class. **Linguistics and Education: An International Research Journal**, v. 17, n. 4, p. 391-406, Inverno de 2006.

CORTES, V.; JONES, J.; STOLLER, F. Lexical bundles in ESP reading and writing. Apresentado na Conferência TESOL, Salt Lake City, Abril 2002.

COUNCIL OF BIOLOGY EDITORS. **Manual de estilo CBE**: Guia para autores, editores y revisores en el campo de la medicina y la biologia. Barcelona: Salvat; 1987.

CRYSTAL, D. A **Dictionary of Linguistics and Phonetics**. Oxford: Blackwell, 3ª ed., p.86. 1991.

DASGUPTA T. Exploiting structural analysis, in silico screening, and serendipity to identify novel inhibitors of drug-resistant falciparum malaria. **ACS Chemical Biology**, EUA, v. 4, n. 1, p. 29-40, 16 jan. 2009.

DAVISSON, M.T. Discovery genetics: serendipity in basic research. **ILAR Journal**, Washington, EUA, v. 46, n. 4, p. 338-45, 2005.

DE COCK, S.; GRANGER, S.; LEECH, G.; MCENERY, T. An automated approach to the phrasicon of EFL learners. In: GRANGER, S. (Org.). **Learner English on computer**. Londres: Longman, 1998. p. 67-79.

DEDALUS Banco de Dados Bibliográficos da USP. Disponível em <<http://dedalus.usp.br:4500/ALEPH/por/USP/USP/DEDALUS/START>>. Último acesso em 03 mar. 2009.

DOAJ (Directory of Open Access Journals). Disponível em <<http://www.doaj.org/>>. Último acesso em 03 mar. 2009.

DÖBEREINER, J. **Especialização e internacionalização de revistas científicas brasileiras** - Objetivo proposto pela CAPES em reunião de Pós-Graduação na USP, em 2006. Disponível em <<http://www.lncc.br/abec/doc/rj/palestras/Dob%20Rio%202007.ppt>>. Acesso 20 out. 2008.

FARIA, L. I. L.; QUONIAM, L.; MUGNAINI, R. Elementos de comparação das bases de dados PASCAL e SCISEARCH. **International Journal of Information Science for Decision Making**, França, Toulon-Var, v. 5, n. 30, dez. 2002.

GOLDENBERG S.; POBLACIÓN D. A.; GOMES P. O.; SOARES A. L.; FERREIRA J. R., KAFEJIAN A. P., BRAGA M. E. R. Editoração de revistas científicas: análise das instruções aos autores de 19 revistas brasileiras. **Acta Cirúrgica Brasileira**, v. 10, n. 2, p. 55-60, 1995.

GRÜNEWALD H. **Guidelines for editors of scientific and technical journals**. Paris: UNESCO/INISIST, 1979.

HASWELL, R. **Gaining ground in college writing**: Tales of development and interpretation. Dallas: Southern Methodist University Press, 1991.

HOEY, M. **Lexical Priming: A New Theory of Words and Language**. London: Routledge, 2005.

HUNSTON, S. **Corpora in Applied Linguistics**. Cambridge: Cambridge University Press, 2002.

JOHNS, T. Micro-Concord: a language learner's research tool. **System**, Oxford, Pergamon, v. 14, n. 2, p. 151-62, 1986.

JUCESP. Deliberação JUCESP no. 004/08 In ATIPIESP, 2009. Disponível em <http://www.atpiesp.org.br/doc/tabela_emolumentos.pdf > Acesso em 20 fev. 2009.

KAPLAN, B. R. **On the scope of Applied Linguistics**. EUA: Newbury House Publishers, 1980.

KENNEDY, G. **An Introduction to Corpus Linguistics**. New York: Longman, 1998.

KENNY, D. Creatures of Habits? What translators usually do with words. **Meta**, v. 43, n. 4, p. 515-23, 1998.

KULAMBIL PADINJAKARA R.N., ASHAWESH K., PATEL V. Allergic reaction to blue cheese: serendipity or actual causation? **The New Zealand Medical Journal**, Nova Zelândia, v. 121, n. 1283, p. 102-4, 3 out. 2008.

LAVIOSA, S. Corpora in translation studies. In BAKER, M. (org) **Routledge Encyclopedia of Translation Studies**. Londres: Routledge, 1998. p. 50-53.

LOPEZ-CÓZAR, E. D. Evaluación del grado de ajuste de las revistas españolas de ciencias de la salud a las normas internacionales de presentación da publicaciones periódicas. **Revista Española de Salud Pública**, Madri, v. 71, n. 6, p. 531-46, nov/dez. 1997.

MAIA, Belinda. Corpora for terminology extraction – the differing perspectives and objectives of researchers, teachers and language services, Portugal, 2002. Disponível em: <www.linguateca.pt/EscolaVerao2006/Corpora/ModuloCorpora.html> . Acesso em: 15 out 2007.

McCULLY, G. Writing quality, coherence, and cohesion, **Research in the Teaching of English**, v. 19, p. 269–82, out. 1985.

MENEZHINI, R. Avaliação da produção científica e o Projeto SciELO. **Ciência da Informação**, Brasília, DF; v. 27, n. 2, p. 219-220, maio/ago. 1998.

MIRANDA D. B.; PEREIRA M. N. F. O periódico científico como veículo de comunicação: uma revisão de literatura. **Ciência da Informação**, Brasília, DF, v. 25, n. 3, p. 375-82, set/dez. 1996.

MOITA LOPES, L. P. **Por uma Linguística Aplicada Indisciplinar**. São Paulo: Parábola, 2006.

MONTEIRO R.; JATENE, F. B.; GOLDENBERG, S.; POBLACIÓN, D. A.; PELLIZZON, R. F. Critérios de autoria em trabalhos científicos: um assunto polêmico e delicado. *Revista Brasileira de Cirurgia Cardiovascular*, São José do Rio Preto, v. 19, n. 4, p. III-VIII, nov/dez. 2004.

MORTON, W.R., SWINDLER K. Serendipitous insights involving nonhuman primates. **ILAR Journal**, Washington, EUA, v. 46, n. 4, p.346-51, 2005.

MUGNAINI, R.; JANNUZZI, P; QUONIAM, L. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Ciência da Informação**, Brasília, DF; v. 33, n. 2, p. 123-131, maio/ago. 2004.

NEWMARK, P. **More Paragraphs on Translation**. Clevedon: Multilingual Matters, 1998.

OTRANTO, M. I. S. **A interação linguageira dentista-paciente na atividade de trabalho de triagem de clínica odontológica**. São Paulo, SP, 2006. Originalmente apresentada como dissertação de mestrado, Pontifícia Universidade Católica de São Paulo, 2006. Disponível em: <http://www.sapientia.pucsp.br/tde_busca/arquivo.php?codArquivo=2177>. Acesso em: 10 jan. 2009.

OWEN, C. *Corpus-Based Grammar and the Heineken Effect: Lexico-Grammatical Description for Language Learners*. **Applied Linguistics**, Oxford, v. 14, n. 2, p. 167-87, 1993.

PARTINGTON, A. Aims, Tools and Practices of *Corpus Linguistics*. **IntUne** University of Bologna, Bologna, jan. 2006. Disponível em: <www.intune.it/file_download/23>. Acesso em 10 jul. 2008.

PEIRCE, C. S. **Collected Papers of Charles Sanders Peirce**, Hartshorne and Weiss (eds.), Cambridge, MA: Harvard University Press, 1932.

PELLIZZON, R.F.; MONTERO E.F.S.; POBLACIÓN, D.A.; MONTEIROS R.; CASTRO, R.C.F. Revistas brasileiras publicadoras de artigos científicos em cirurgia. III: análise das instruções aos autores baseada na estrutura dos requisitos de Vancouver. **Acta Cirúrgica Brasileira**, v. 22, n. 6, p. 503-10, Nov.-Dez. 2007.

PEREZ GÓMEZ, M. A. Papel del editor y de los comites editoriales como guardianes de la calidad de las revistas. **Ciencias de la Información**, La Habana v. 29, p. 37-44, 1998.

PERROTTI-GARCIA, A. J. Análise da Classificação de Biber para *Clusters* de um *Corpus* de Textos Médicos de Autores Brasileiros: Estudo Piloto. In: 56º. SEMINÁRIO DO GEL, 2008, São José do Rio Preto. **Resumos**. São José do Rio Preto: Grupo de Estudos Linguísticos, 2008.

PERROTTI-GARCIA, A.J. **Customised corpora – a source of information for translators**. In: INSTITUTE OF TRANSLATION AND INTERPRETING CONFERENCE 2007, 21-22 de abril de 2007, Londres, Inglaterra. Disponível em: <<http://www.iti-conference.org.uk/itic21/index.html>>. Acesso 15 nov. 2008.

PERROTTI-GARCIA, A.J. (2005) “O Uso de *Corpus* Customizado como Fonte de Pesquisa para Tradutores”, **Confluências Revista de Tradução Científica e Técnica**, Lisboa, v. 3, p. 62-79. Disponível em: <<http://www.confluencias.net/>>. Acesso em 20 mar. 2008.

POBLACIÓN, D.A.; GOLDENBERG, S.; MONTERO E.F.S.; GOMES, P. O.; CAMPANATTI-OSTIZ, H.; MOREIRA, M. B.; SANTOS, A. C.; PELLIZZON, R. F. Revistas brasileiras publicadoras de artigos científicos em cirurgia. II - Terminologia e atribuições adotadas pelos editores. Proposta de organograma do periódico e fluxograma do artigo. **Acta Cirúrgica Brasileira**, v. 18, n. 6, p. 497-501, Nov.-Dez. 2003.

POBLACIÓN, D.A.; GOLDENBERG, S.; MONTERO E.F.S.; MOREIRA, M.B.; PELLIZZON, R.F.; Revistas brasileiras publicadoras de artigos científicos em cirurgia. I: características estruturais e administrativas das revistas. **Acta Cirúrgica Brasileira**, v. 17, n. 6, p. 359-69, Nov.-Dez. 2002.

PubMed Base de dados da National Library of Medicine. Disponível em <<http://www.ncbi.nlm.nih.gov/pubmed/>>. Último acesso em 03 mar. 2009.

RELMAN A.S. Publish or perish - or both. **New England Journal of Medicine**, Massachusetts, v. 297, n. 13, p. 724-5, set. 1977.

RIBEIRO, M.P.F. A importância da indexação para a difusão do conhecimento comunicado nas revistas técnico-científicas. **Revista Mineira de Enfermagem**, v. 10, n. 1 (Editorial) Jan. 2006. Disponível em <http://www.portalbvsenf.eerp.usp.br/scielo.php?script=sci_arttext&pid=S1415-27622006000100001&lng=pt&nrm=iso&tlng=pt> Acesso 20 fev. 2009.

SAALFRANK, R.W.; MAID H.; SCHEURER A. Supramolecular coordination chemistry: the synergistic effect of serendipity and rational design. **Angewandt Chemie** (International Edition in English), Alemanha, v. 47, n. 46, p. 8794-824, 2008.

SARMENTO E SOUZA M. F.; GREGÓRIO, S. A. B.; FORETI, M. C. P. P. Critérios de qualidade em artigos e periódicos científicos: da mídia impressa à eletrônica. **Revista Transformação**, Campinas, v.16, n. 1, p. 71-89, jan/abril 2004.

SCHLEPPEGRELL, M. J. Conjunction in spoken English and ESL writing. **Applied Linguistics**, Oxford, v. 17, n. 3, p. 271-85, 1996.

SCHRIGER D. A.; WEARS R. L.; COOPER R. J.; CALLAHAM M. L. Upgrading our instructions for authors. **Annals of Internal Medicine**, v. 41, n. 4, p. 565-7, 2003.

SciELO Scientific Electronic Library Online. Disponível em <<http://www.scielo.br/scielo.php?lng=pt>>. Último acesso em 03 mar. 2009.

SCOTT, M. **WordSmith Tools version 4**, Oxford: Oxford University Press, 2004.

SCOTT, M. PC Analysis of Key Words -- and Key Key Words, **System**, v. 25, n. 1, p. 1-13, 1997.

SCOTT, M.; TRIBBLE C. English for academic purposes. **Textual Patterns. Key words and corpus analysis in language education**. Philadelphia: John Benjamins Publishing Company, 2006. p. 131- 160.

SERRALVO, G. **Processo de versão de um artigo acadêmico de marketing com o auxílio de corpora eletrônicos e da web**. São Paulo, SP, 2004. Originalmente apresentada como

dissertação de mestrado, Pontifícia Universidade Católica de São Paulo, 2004. Disponível em: <http://www.pucsp.br/pos/lael/lael-inf/def_teses.html>. Acesso em: 10 jan. 2009.

SERRANO, Daniel Portillo. Serendipite ou Serendipity. 20 set.2007. Disponível em <http://www.portaldomarketing.com.br/Artigos/Serendipite_ou_Serendipity.htm>. Acesso em 14 dez. 2008.

SHERGUE, O. **Dimensão de Variação no Discurso Médico Acadêmico**: o artigo de pesquisa e a apresentação de trabalhos científicos em congressos. São Paulo, SP. Originalmente apresentada como dissertação de mestrado, Pontifícia Universidade Católica de São Paulo, 2003. Disponível em: <http://www.pucsp.br/pos/lael/lael-inf/def_teses.html>. Acesso em: 10 jan. 2009.

SIGNORINI, I.; CAVALCANTI, M. (orgs.). **Linguística Aplicada e transdisciplinaridade: questões e perspectivas**. Campinas: Mercado das Letras, 2004.

SINCLAIR, J. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.

SINCLAIR, J. **English Collocation Studies: the OSTI report**. Continuum International Publishing Group, 2004.

SINTRA. **Repensando o curriculum de um curso superior em tradução**. Rio de Janeiro, sem data especificada. Disponível em <<http://www.sintra.org.br/site/index.php?p=c&id=14&codcat=13>>. Acesso em 20 fev. 2009.

SORINOLA, O.; OLUFOWOBI, O.; COOMARASAMY, A.; KHAN, K. S. Instructions to authors for case reporting are limited: a review of a core journal list. **BMC Medical Education**, v. 4, n. 4, 2004.

SOSA-DE-MARTINEZ, M. C.; CARNEVALE, A.; AVENDAÑO-INESTRILLAS, J. Comparación de las normas para autores de revistas médicas mexicanas con las normas de Vancouver. **Revista de Investigación Clínica**, México, v. 47, n. 3, p. 203-10, 1995.

SPINAK, E. Las analisis cuantitativos de la literatura científica y su validez para juzgar la producción latinoamericana. **Boletín de la Oficina Sanitaria Panamericana**, v. 120, n. 2, p. 139-47, 1996.

STEVENSON, A.; TUOHY C.; NORRISH, J. **Ambient Findability and Structured Serendipity: Enhanced Resource Discovery for Full Text Collections**. Auckland: Victoria University of Wellington, 2008.

STOSKOPF, M. K. Observation and cogitation: how serendipity provides the building blocks of scientific discovery. **ILAR Journal**, Washington, EUA, v. 46, n. 4, p. 332-7, 2005.

STUBBS, M. Language Corpora. In: DAVIES, A. & ELDER, C. **The handbook of applied linguistics**. Oxford: Blackwell Publishing, 2004. p. 106-132.

SUSSMAN, R. Serendipity: Reflections of my experience in the scientific world. **The Journal of Biological Chemistry**, EUA, 15 dez. 2008. Disponível em <<http://www.jbc.org/cgi/reprint/X800013200v1>> Acesso 15 fev. 2009.

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

UNAERP. **Currículo 652 / 4C Lic. em Letras (Port./Inglês)**. Ribeirão Preto, 2008. Disponível em www.unaerp.br/index2.php?option=com_content&do_pdf=1&id=403 >. Acesso em 20 fev. 2009.

UNIP. **Grade Curricular - Letras** (Licenciatura em Línguas Portuguesa/Inglesa e Bacharelado em Tradução). São Paulo, 2008. Disponível em <http://www3.unip.br/ensino/graduacao/tradicionais/hum_letras_grade.aspx >. Acesso em 25 jan. 2009.

VIEIRA, M. A. M. **A atividade, o discurso e a clínica. Uma análise dialógica do trabalho médico**. São Paulo, SP. Originalmente apresentada como tese de doutorado, Pontifícia Universidade Católica de São Paulo, 2002. Disponível em: <http://www.pucsp.br/pos/lael/lael-inf/def_teses.html>. Acesso em: 10 jan. 2009.

WETZEL, L. Types and Tokens. In: ZALTA E. N. (ed.). **The Stanford Encyclopedia of Philosophy** (Winter 2007 ed.). Disponível em: <<http://plato.stanford.edu/archives/win2007/entries/types-tokens/>>. Acesso em 14 mar. 2008.

WILKINSON, M. Corpora, Serendipity & Advanced Search Techniques. **The Journal of Specialised Translation**, v. 7, p.108-122. Jan, 2007. Disponível em <http://www.jostrans.org/issue07/art_wilkinson.pdf>. Acesso 10 jun.2008.

WORMELL I. Informetria: explorando bases de dados como instrumentos de análise. **Ciencias de la Información**, La Habana v. 27, n. 2, p. 210-6, 1998.

APÊNDICES

APÊNDICE A: Periódicos dos quais foram coletados artigos para o *corpus* de estudo, com sua respectiva entidade publicadora e classificação QUALIS.

Título do Periódico	Entidade Publicadora	QUALIS
Revista Brasileira de Anestesiologia	Sociedade Brasileira de Anestesiologia	A/B3
<i>São Paulo Medical Journal</i>	Associação Paulista de Medicina	B3
Jornal de Pediatria	Sociedade Brasileira de Pediatria	B3
<i>Brazilian Dental Journal</i>	Fundação Odontológica de Ribeirão Preto	B3
Jornal Vascular Brasileiro	Sociedade Brasileira de Angiologia e de Cirurgia Vascular	B5
Revista Brasileira de Medicina do Esporte	Sociedade Brasileira de Medicina do Esporte	B5
<i>Brazilian Journal of Infectious Diseases</i>	Sociedade Brasileira de Doenças Infecciosas	B2
Revista do Instituto de Medicina Tropical de São Paulo	Instituto de Medicina Tropical de São Paulo	B3
Revista Brasileira de Psiquiatria	Associação Brasileira de Psiquiatria	B2
<i>International Brazilian Journal of Urology</i>	Sociedade Brasileira de Urologia	B3
Arquivos Brasileiros de Endocrinologia e Metabologia	Sociedade Brasileira de Endocrinologia e Metabologia	B3
Revista Brasileira de ORL	Associação Brasileira de Otorrinolaringologia e Cirurgia Cérvico-Facial	B3
<i>Clinics</i>	Faculdade de Medicina da USP	B5
<i>Brazilian Journal of Medical and Biological Research</i>	Associação Brasileira de Divulgação Científica, composta pelas seguintes entidades científicas: Sociedade Brasileira de Biofísica; Sociedade Brasileira de Farmacologia e Terapêutica Experimental; Sociedade Brasileira de Fisiologia; Sociedade Brasileira de Imunologia; Sociedade Brasileira de Investigação Clínica; Sociedade Brasileira de Neurociências e Comportamento.	B2

APÊNDICE B: Periódicos dos quais foram coletados artigos para o *corpus* de referência, com sua respectiva entidade publicadora.

Título do Periódico	Entidade Publicadora	Tipo de revisão
Revistas publicadas nos Estados Unidos		
<i>Pediatrics</i>	<i>American Academy of Pediatrics</i>	Revisão por pares
<i>Clinical Diabetes</i>	<i>American Diabetes Association</i>	Conselho editorial
<i>American Journal of Otolaryngology: Head Neck Medicine and Surgery</i>	<i>Elsevier Inc.</i>	Conselho editorial
<i>American Journal of Cardiology</i>	<i>Excerpta Medica</i>	Revisão por pares
<i>The Internet Journal of Third World Medicine</i>	<i>American College of Cardiology</i>	Revisão por pares
<i>The Internet Journal of Emergency Medicine</i>	<i>Internet Scientific Publications LLC</i>	Revisão por pares
<i>American Journal of Nephrology</i>	<i>Internet Scientific Publications LLC</i>	Revisão por pares
<i>American Journal of Clinical Oncology</i>	<i>Karger AG, Basel</i>	Revisão por pares
	<i>Lippincott Williams & Wilkins</i>	Revisão por pares
	<i>American College of Radiation</i>	

	<i>Oncology</i>	
	<i>Lippincott Williams & Wilkins</i>	Revisão por pares
<i>Stroke</i>		
<i>Clinical Medicine & Research</i>	<i>Marshfield Clinic</i>	Revisão por pares
Revistas publicadas na Europa		
<i>British Medical Journal</i>	BMJ Group	Revisão por pares
<i>British Medical Journal of Ophthalmology</i>	BMJ Group	Revisão por pares
<i>British Journal of Ophthalmology</i>	BMJ Group	Revisão por pares
<i>BMC</i>	BioMed Central	Revisão por pares
<i>Journal of Vascular Research</i>	Karger AG, Basel	Revisão por pares
<i>The Lancet</i>	Elsevier Inc	Revisão por pares
<i>Geriatrics</i>	Modern Medicine	Revisão por pares
<i>Orphanet Journal of Rare Diseases</i>	BioMed Central	Revisão por pares
	BioMed Central	Revisão por pares
<i>Journal of Neuroinflammation</i>		
<i>BioPsychoSocial Medicine-</i>	BioMed Central	Revisão por pares
<i>International Journal of Behavioral</i>	BioMed Central	Revisão por pares
<i>Nutrition and Physical Activity</i>		
<i>Blood Disorders</i>	BioMed Central	Revisão por pares
<i>Endocrine Disorders</i>	BioMed Central	Revisão por pares
<i>Nuclear Medicine (UK)</i>	BioMed Central	Revisão por pares
<i>AIDS Research and Therapy</i>	BioMed Central	Revisão por pares
<i>Head and Face Medicine</i>	BioMed Central	Revisão por pares
<i>Emergency Medicine</i>	BioMed Central	Revisão por pares
<i>Immunity and Ageing</i>	BioMed Central	Revisão por pares
<i>Pediatric Rheumatology-</i>	BioMed Central	Revisão por pares
<i>Journal of Occupational Medicine and</i>	BioMed Central	Revisão por pares
<i>Toxicology</i>		
<i>Cough</i>	BioMed Central	Revisão por pares

APÊNDICE C - Rotina em linguagem Shell, para a determinação dos trigramas exclusivos do *corpus* BRAZIL (desenvolvido por Ana Julia Perrotti-Garcia)

```
#3BRAZIL.txt
#3ABROAD.txt
cat 3BRAZIL.txt 3ABROAD.txt 3ABROAD.txt | sort | uniq -c | grep 1 >
temsonobrazil.txt
echo Arquivo criado com sucesso
```

APÊNDICE D - Trigramas mais frequentes no *corpus* de referência (ABROAD), totalizando 85 trigramas com frequência igual ou superior a 0,01%.

ordem	trigrama	frequência	%
1	in patients with	2.696	0,05
2	context link #	1.953	0,03
3	in a new	1.921	0,03

4	a new window	1.907	0,03
5	window in a	1.907	0,03
6	in this window	1.906	0,03
7	this window in	1.906	0,03
8	as well as	1.849	0,03
9	the presence of	1.585	0,03
10	links context link	1.574	0,03
11	methods results discussion	1.573	0,03
12	bibliographic links context	1.572	0,03
13	the use of	1.490	0,03
14	of the #	1.482	0,03
15	the number of	1.362	0,02
16	in this study	1.276	0,02
17	of patients with	1.191	0,02
18	in table #	1.138	0,02
19	the present study	1.125	0,02
20	view this table	1.115	0,02
21	results discussion references	1.082	0,02
22	top abstract introduction	1.057	0,02
23	of the study	961	0,02
24	the effect of	915	0,02
25	the risk of	914	0,02
26	type # diabetes	888	0,02
27	there was no	843	0,02
28	from # to	840	0,01
29	the development of	836	0,01
30	according to the	812	0,01
31	one of the	807	0,01
32	the time of	802	0,01
33	larger version #k	791	0,01
34	view larger version	791	0,01
35	of # patients	779	0,01
36	help with image	776	0,01
37	image viewing email	776	0,01
38	viewing email	776	0,01

	jumpstart		
39	with image viewing	776	0,01
40	email jumpstart to	775	0,01
41	jumpstart to image	775	0,01
42	version of this	769	0,01
43	in # of	766	0,01
44	more likely to	766	0,01
45	at the time	762	0,01
46	of this study	753	0,01
47	at least #	733	0,01
48	this study was	728	0,01
49	was used to	728	0,01
50	years of age	725	0,01
51	the effects of	723	0,01
52	based on the	721	0,01
53	was associated with	712	0,01
54	between # and	707	0,01
55	of the patients	702	0,01
56	in the present	694	0,01
57	in # patients	693	0,01
58	the prevalence of	690	0,01
59	a total of	672	0,01
60	is associated with	666	0,01
61	in our study	665	0,01
62	total of #	661	0,01
63	been shown to	653	0,01
64	in the study	652	0,01
65	the results of	648	0,01
66	the treatment of	646	0,01
67	new window table	635	0,01
68	window table #	625	0,01
69	because of the	619	0,01
70	top abstract methods	617	0,01
71	abstract methods	616	0,01
	results		
72	materials and	616	0,01
	methods		
73	in order to	605	0,01
74	in addition to	604	0,01

75	a number of	597	0,01
76	up to #	589	0,01
77	was defined as	588	0,01
78	and can be	586	0,01
79	associated with a	581	0,01
80	as a result	577	0,01
81	compared with the	570	0,01
82	the absence of	569	0,01
83	the united states	567	0,01
84	there were no	567	0,01
85	to # years	563	0,01

APÊNDICE E - Trigramas mais frequentes no *corpus* de estudo (BRAZIL), totalizando 111 trigramas com frequência igual ou superior a 0,01%.

ordem	trigrama	frequência	%
1	p #	6.686	0,12
2	the use of	2.893	0,05
3	the presence of	2.689	0,05
4	the present study	2.639	0,05
5	in order to	2.267	0,04
6	as well as	2.179	0,04
7	n #	1.855	0,03
8	according to the	1.831	0,03
9	in patients with	1.801	0,03
10	in the present	1.662	0,03
11	the number of	1.579	0,03
12	version of this	1.541	0,03
13	larger version of	1.540	0,03
14	view larger version	1.540	0,03
15	k jpg file	1.531	0,03
16	due to the	1.522	0,03
17	from # to	1.514	0,03
18	jpg file	1.446	0,03
19	in this study	1.409	0,03
20	one of the	1.377	0,03
21	between # and	1.302	0,02
22	of the patients	1.213	0,02
23	the development of	1.191	0,02

24	of the #	1.109	0,02
25	the treatment of	1.104	0,02
26	in # of	1.052	0,02
27	of patients with	1.040	0,02
28	there was no	1.035	0,02
29	of this study	1.029	0,02
30	related to the	1.013	0,02
31	the prevalence of	1.004	0,02
32	to # years	1.000	0,02
33	on the other	994	0,02
34	of this image	987	0,02
35	the other hand	980	0,02
36	this image #	974	0,02
37	image # k	973	0,02
38	based on the	957	0,02
39	the end of	938	0,02
40	the control group	923	0,02
41	the results of	923	0,02
42	in table #	920	0,02
43	in the study	917	0,02
44	this study was	901	0,02
45	in the literature	900	0,02
46	more than #	898	0,02
47	study was to	896	0,02
48	less than #	891	0,02
49	there was a	884	0,02
50	it has been	881	0,02
51	the fact that	827	0,02
52	in relation to	824	0,02
53	de são paulo	818	0,02
54	of são paulo	814	0,02
55	there is a	812	0,02
56	the incidence of	810	0,02
57	an increase in	808	0,02
58	of the disease	804	0,01
59	increase in the	798	0,01
60	compared to the	797	0,01
61	the risk of	795	0,01
62	of the present	770	0,01

63	in our study	743	0,01
64	up to #	740	0,01
65	of # patients	731	0,01
66	the effects of	723	0,01
67	the diagnosis of	722	0,01
68	to evaluate the	716	0,01
69	associated with the	704	0,01
70	material and methods	704	0,01
71	for # min	699	0,01
72	the effect of	698	0,01
73	in addition to	693	0,01
74	the time of	672	0,01
75	analysis of the	669	0,01
76	was used to	668	0,01
77	by means of	660	0,01
78	considered to be	654	0,01
79	of the study	651	0,01
80	in # patients	637	0,01
81	when compared to	637	0,01
82	the most common	636	0,01
83	to # of	632	0,01
84	end of the	628	0,01
85	in # #	623	0,01
86	quality of life	620	0,01
87	the absence of	618	0,01
88	in the first	614	0,01
89	approved by the	613	0,01
90	the aim of	609	0,01
91	to determine the	607	0,01
92	the level of	603	0,01
93	in terms of	595	0,01
94	observed in the	594	0,01
95	years of age	588	0,01
96	the importance of	587	0,01
97	total of #	585	0,01
98	at the end	584	0,01
99	found in the	584	0,01
100	is associated with	583	0,01

101	and in the	581	0,01
102	at least #	578	0,01
103	a total of	574	0,01
104	rio de janeiro	573	0,01
105	there were no	572	0,01
106	there is no	567	0,01
107	the majority of	556	0,01
108	of this table	553	0,01
109	at the time	547	0,01
110	and # #	546	0,01
111	such as the	546	0,01

Apêndice F - Trigramas exclusivos do *corpus* BRAZIL com ocorrência igual ou superior a 100, totalizando 185 trigramas (dos 2.504 trigramas exclusivos encontrados).

Trigrama	Número de ocorrências
LARGER VERSION OF	1.540
K JPG FILE	1.531
JPG FILE	1.446
THIS IMAGE #	974
IMAGE # K	973
DE SÃO PAULO	818
OF SÃO PAULO	814
RIO DE JANEIRO	573
OF THIS TABLE	553
THIS TABLE #	537
TABLE # K	537
FILE TABLE	514
CORRESPONDENCE ABSTRACT	502
UNIVERSITY OF SÃO	500
WERE SUBMITTED TO	460
UNIVERSIDADE FEDERAL DE	445
FACULDADE DE MEDICINA	425
ADDRESS FOR CORRESPONDENCE	421
CORRESPONDENCE AND FOOTNOTES	418
ARE REPORTED AS	417
AND FOOTNOTES	417
FOOTNOTES ABSTRACT	414
UNIVERSIDADE DE SÃO	408
FILE FIGURE	399
FEDERAL UNIVERSITY OF	396
FOR CORRESPONDENCE	345
DATA ARE REPORTED	319
HOSPITAL DAS CLÍNICAS	312
OBSERVED THAT THE	280
FILE DISCUSSION	279
BRAZIL ADDRESS FOR	278
THE CITY OF	266
FEDERAL DE SÃO	261
AT #FC FOR	244

THE RESULTS OBTAINED	241
BRASIL ABSTRACT INTRODUCTION	238
IT WAS OBSERVED	233
UNIVERSIDADE FEDERAL DO	227
INCREASE OF THE	227
REFERENCES CORRESPONDENCE AND	226
DID NOT PRESENT	226
REPORTED AS MEANS	224
INTRODUCTION MATERIAL AND	219
ABSTRACT INTRODUCTION MATERIAL	219
PEDIATR RIO J	216
DUE TO ITS	215
RIO GRANDE DO	214
IN THE CITY	213
OBJECTIVE OF THE	206
OF THE BRAZILIAN	202
WE OBSERVED THAT	199
IN SÃO PAULO	198
UNIVERSITY MEDICAL SCHOOL	195
THE CHI-SQUARE	195
GRANDE DO SUL	195
ACKNOWLEDGMENTS CORRESPONDENCE AND	189
REFERENCES ACKNOWLEDGMENTS CORRESPONDENCE	188
LOWER THAN #	188
SÃO PAULO UNIVERSITY	186
STUDENT T-TEST	184
ALL OF THEM	181
THE RESEARCH ETHICS	178
PAULO UNIVERSITY MEDICAL	172
WITH THE AIM	170
SUBMITTED TO THE	168
WAS OBSERVED THAT	167
DISCUSSION REFERENCES CORRESPONDENCE	167
THE #TH DAY	166
PATIENTS SUBMITTED TO	166
WAS SUBMITTED TO	162
OF THE TEST	160
DETERMINATION OF THE	160
OF RIBEIRÃO PRETO	158
IN THE STATE	158
DE MEDICINA DE	157
AUTHORS CONCLUDED THAT	157
SUBMITTED TO A	155
AGED BETWEEN #	155
MANN-WHITNEY TEST	151
VARIED FROM #	149
FEDERAL DO RIO	149
THE NEED OF	148
STATE UNIVERSITY OF	146
THE PRESENT INVESTIGATION	144
PAULISTA DE MEDICINA	144
OF THE VACCINE	143
OF THE HOSPITAL	143
CHRONIC HEPATITIS C	143
WAS OF #	142
OF THE INDIVIDUALS	142
ON THE #TH	141
SANTA CASA DE	140

THE FEDERAL UNIVERSITY	139
THE AUTHORS CONCLUDED	138
A PREVALENCE OF	137
DISCUSSION REFERENCES ACKNOWLEDGMENTS	136
ESCOLA PAULISTA DE	134
DE RIBEIRÃO PRETO	134
USE OF ANTIBIOTICS	133
OF DENTISTRY OF	133
WHEN THERE IS	132
ON TABLE #	132
IN RIBEIRÃO PRETO	132
THE LOWER LIMBS	131
STUDY CARRIED OUT	131
ORDER TO DETERMINE	130
MIN AT #FC	130
DETECTED IN #	130
UNIVERSIDADE ESTADUAL DE	129
OR EQUAL TO	129
AT -#FC	129
SINCE IT IS	128
SCHOOL OF DENTISTRY	128
OF THE SURGICAL	128
BRAZIL CORRESPONDENCE	128
ORDER TO AVOID	126
AGREEMENT WITH THE	126
UNDER # YEARS	125
TO VERIFY THE	125
MG KG-#	124
THE BEHAVIOR OF	122
SET AT P	122
THE MOST PREVALENT	121
INTENSITY OF THE	121
OF RIO DE	120
MEANS OF A	120
INFORMED CONSENT FORM	120
THE SPECIMENS WERE	119
THE PARTICIPATION OF	119
OF HEARING LOSS	119
DIFFERENCE WAS OBSERVED	119
LENGTH OF TIME	118
BRASIL #DEPARTAMENTO DE	118
KG-# DAY	117
IN ## #	117
FREQUENCY OF THE	117
STATE OF SÃO	116
OF CHOICE FOR	116
THE HPA AXIS	114
THE GROUP WITH	114
THE POSTOPERATIVE PERIOD	113
OF THE FEDERAL	113
IT WAS POSSIBLE	113
THE STUDENT T	112
HEPATITIS C VIRUS	112
THE PRESENT RESULTS	111
DUE TO THEIR	111
DEPARTMENT OF UROLOGY	111
WAS POSSIBLE TO	110
ONE OF THEM	110

A PREDOMINANCE OF	110
PAULO MEDICAL SCHOOL	109
IT IS BELIEVED	109
DO RIO GRANDE	109
DE PORTO ALEGRE	109
HAS ALREADY BEEN	108
ABSTRACT TEXT REFERENCES	108
INTENSIVE CARE UNITS	107
TO PERFORM THE	106
THREE OR MORE	106
SOME AUTHORS HAVE	106
MOST OF THEM	106
DE MEDICINA DA	106
MALE WISTAR RATS	105
INTERFERE WITH THE	105
ESTADUAL DE CAMPINAS	105
DO RIO DE	105
CAN BE OBSERVED	105
PRODUCED BY THE	104
MEDICINA DE RIBEIRÃO	104
DE CIÊNCIAS MÉDICAS	104
CITY OF SÃO	104
LARGER NUMBER OF	103
IN OUR COUNTRY	103
AND METHODS THE	103
THE VALUES OF	102
THAN # DAYS	102
CRITICALLY ILL PATIENTS	102
WITH THE PURPOSE	101
THE UNDERSTANDING OF	101
THE RESULTS SHOWED	101
THE CASES OF	101
SPIKE OF THE	101
AFTER THE END	101
CHILDREN UNDER #	100

APÊNDICE G: Trigramas exclusivos do *corpus* BRAZIL; em ordem alfabética em sem as frequências de ocorrência, primeiros 25 resultados obtidos. # = números do texto original, convertidos pelo programa no símbolo cerquilha ou jogo da velha.

Trigramas exclusivos do <i>corpus</i> BRAZIL (primeiros 25 resultados)
A # CM
A # CONFIDENCE
A # TO
A # X
A #-BP
A #-H
A #-MIN

A CASE-CONTROL
 A DECREASE OF
 A DELAY IN
 A FACT THAT
 A FINAL VOLUME
 A FREQUENCY OF
 A GREATER NUMBER
 A GREATER RISK
 A HIGH FREQUENCY
 A HIGH INCIDENCE
 A HIGH PREVALENCE
 A HIGH RATE
 A HIGH-FAT
 A HIGHER FREQUENCY
 A HIGHER NUMBER
 A LARGER NUMBER
 A LONG PERIOD
 A LONG TIME

APÊNDICE H: Rotina em linguagem Shell, para a recuperação das frequências dos trigramas exclusivos do *corpus* BRAZIL (desenvolvida pelo Prof. Dr. Tony Berber Sardinha)

```

#check if files are true ascii
#head *.txt | less

#4EUROPA.txt is binary! convert with col
#col -b < 4EUROPA.txt > a
#mv a 4EUROPA.txt

while read f1 f2
do
echo "doing $f1 and $f2....."
sed 's/[ ]*/' $f1 | tr ' ' '_' | gcut -d '_' -f2- | gsort | tr -d '\r' > a
tr ' ' '_' < $f2 | tail +2 | gsort -k2 | tr -d '\r' > b
gjoin -1 1 -2 2 -o 1.1 -o 2.3 a b | tr -d '\.' | sort -nr -k2 | grep '[0-9]' > $f1.freqs
done < files_to_compare

```

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)